# NAVAL POSTGRADUATE SCHOOL

**MONTEREY, CALIFORNIA**

# THESIS

**AN ANALYSIS OF TESTING RISKS:**
**A STRATEGY FOR MITIGATION**

by

Al Mousseau

September 2004

| | |
|---|---|
| Thesis Advisor: | John Dillard |
| Second Reader: | Walter Owen |
| Associate Advisor: | Scott Weed |

**Approved for public release; distribution is unlimited**

THIS PAGE INTENTIONALLY LEFT BLANK

**REPORT DOCUMENTATION PAGE** | *Form Approved OMB No. 0704-0188*



| | | |
|---|---|---|
| **1. AGENCY USE ONLY *(Leave blank)*** | **2. REPORT DATE** September 2004 | **3. REPORT TYPE AND DATES COVERED** Master's Thesis |

**4. TITLE AND SUBTITLE**: An Analysis of Testing Risks: A Strategy for Mitigation

**5. FUNDING NUMBERS**

**6. AUTHOR(S) LCDR Al Mousseau, Jr**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Naval Postgraduate School
Monterey, CA 93943-5000

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
N/A

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES** The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the US Government.

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Distribution Statement (mix case letters)

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (maximum 200 words)**

The Department of Defense is continuously trying to improve the product development effort for its weapon systems. As the complexity of those systems increases, so does the importance of the test and evaluation process. All Services have been victims of poor performance in the independent Operational Evaluation of their respective weapon systems. With the drive to deliver products rapidly to the Warfighter, the prospect for success in Operational Test is reduced. Years of neglect and funding reductions have resulted in a decaying test infrastructure. The acquisition community's failure to consistently apply lessons learned and best business practices ensures repeating the mistakes. The US Navy is embarking on an aggressive six-year development effort to retrofit the aging High speed Anti-Radiation Missile with advanced technology and net-centric enabling systems. This Sea Power 21 weapon will require a test strategy that can effectively verify and evaluate product maturity before independent operational testing. Applying best business practices, lessons learned, and understanding the current state of affairs with respect to the range infrastructure, the Advanced Anti-Radiation Guided Missile Test and Evaluation Integrated Product Team can develop a test approach to mitigate the risk of operational test failure.

**14. SUBJECT TERMS**
Test and Evaluation, systems engineering, AARGM weapon system, range infrastructure, lessons learned

**15. NUMBER OF PAGES** 139

**16. PRICE CODE**

| **17. SECURITY CLASSIFICATION OF REPORT** | **18. SECURITY CLASSIFICATION OF THIS PAGE** | **19. SECURITY CLASSIFICATION OF ABSTRACT** | **20. LIMITATION OF ABSTRACT** |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release; distribution is unlimited**

# AN ANALYSIS OF TESTING RISKS: A STRATEGY FOR MITIGATION

Albert G. Mousseau, Jr.
Lieutenant Commander, United States Navy
B.S., Worcester Polytechnic Institute, 1990

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN SYSTEMS ENGINEERING**

from the

**NAVAL POSTGRADUATE SCHOOL**
**September 2004**

Author: LCDR Al Mousseau, Jr.

Approved by: John Dillard
Thesis Advisor

Walter E. Owen, DPA
Second Reader

Scott Weed
Associate Advisor

Phil DePoy, Ph.D.
Director, Wayne E. Meyer Institute of Systems Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

The Department of Defense continuously seeks to improve the product development effort for its weapon systems. As the complexity of those systems increases, so does the importance of the test and evaluation process. All Services are victims of poor performance in the independent Operational Evaluation of their respective weapon systems. The drive to deliver products rapidly to the Warfighter reduces the prospect for success in Operational Test. Years of neglect and funding reductions have resulted in a decaying test infrastructure. The acquisition community's failure to consistently apply lessons learned and best business practices ensures repeating the mistakes. The US Navy embarked on an aggressive six-year development effort to retrofit the aging High-speed Anti-Radiation Missile with advanced technology and net-centric enabling systems. This Sea Power 21 weapon requires a test strategy that can effectively verify and evaluate product maturity before independent operational testing. By applying best business practices, lessons learned, and understanding the current state of affairs with respect to the range infrastructure, the Advanced Anti-Radiation Guided Missile Test and Evaluation Integrated Product Team can develop a test approach to mitigate the risk of operational test failure.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

# LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| A/A | – Air-to-Air |
| AARGM | – Advance Anti-Radiation Guided Missile |
| ACAT | – Acquisition Category |
| ACTD | – Advanced Concept Technology Demonstration |
| ADATS | – Air Defense Anti-Tank System |
| ATD | – Advanced Technology Demonstration |
| ARH | – Anti-Radiation Homing |
| AWL | – Advanced Weapons Lab |
| BRAC | – Base Realignment and Closure |
| CDD | – Capabilities Description Document |
| COMOPTEVFOR | – Commander Operational Test Force |
| COI | – Critical Operational Issue |
| CONOPS | – Concept of Operations |
| CPD | – Capabilities Production Document |
| DoD | – Department of Defense |
| DoDD | – DoD Directive |
| DoDI | – DoD Instruction |
| DOT&E | – Director, Operational Test & Evaluation |
| DSB | – Defense Science Board |
| DT | – Developmental Test |
| DT&E | – Developmental Test & Evaluation |
| DTRMC | – DoD Test and Resource Management Center |
| EMCON | – Emissions Control |
| FOT&E | – Follow-on Test and Evaluation |
| FNC | – Future Naval Capability |
| GAO | – Government Accounting Office |
| HARM | – High-Speed Anti-Radiation Missile |
| IADS | – Integrated Air Defense System |
| ICD | – Initial Capabilities Document |
| ILS | – Integrated Logistics Support |
| IOC | – Initial Operational Capability |
| IOT&E | – Independent Operational Test & Evaluation |
| IPT | – Integrated Product Team |
| IT&E | – Integrated Test & Evaluation |

| | |
|---|---|
| JROC | – Joint Requirements Oversight Council |
| JSEAD | – Joint Suppression of Enemy Air Defenses |
| KPP | – Key Performance Parameter |
| LRIP | – Low-Rate Initial Production |
| MRTFB | – Major Range and Test Facility Base |
| MDA | – Missile Defense Agency |
| MMW | – Millimeter Wave |
| MNS | – Mission Needs Statement |
| MOE | – Measure of Effectiveness |
| MOS | – Measure of Suitability |
| NASA | – National Aeronautics Space Administration |
| NDIA | – National Defense Industrial Association |
| ONR | – Office of Naval Research |
| OPEVAL | – Operational Evaluation |
| ORD | – Operational Requirements Document |
| OT | – Operational Test |
| OTG | –Operational Test Guide |
| PEO | – Program Executive Officer |
| PM | – Program Manager |
| PMO | – Program Management Office |
| PRM | – Production Representative Missile |
| RCS | – Radar Cross Section |
| RF | – Radio Frequency |
| SAM | – Surface to Air Missile |
| SAMP | – System Acquisition Master Plan |
| SEAD | – Suppression of Enemy Air Defenses |
| SD&D | – System Design & Development |
| SLAM | – Standoff Land Attack Missile |
| SLAM-ER | – SLAM Expanded Response |
| SOAD | – Standoff Outside of Area Defense |
| T&E | – Test & Evaluation |
| TEMP | – Test and Evaluation Master Plan |
| THAAD | – Theater High Altitude Area Defense |
| TPWG | – Test Plan Working Group |
| TRL | – Technology Readiness Levels |
| TSSAM | – Tri-Service Standoff Attack Missile |
| TWG | – Targets Working Group |

| | |
|---|---|
| USD(AT&L) | – Under Secretary of Defense (Acquisition, Technology & Logistics) |
| VV&A | – Verification, Validation & Accreditation |
| WIA | – Weapons Impact Assessment |

THIS PAGE INTENTIONALLY LEFT BLANK

# ACKNOWLEDGMENTS

I am in debt to Lori, Matt and Ashley. Their sacrifice through these two years has been beyond what should be asked of an already sacrificing military family.

I would like to thank Mr. Gordon Turner and Mr. Ken Claunch for their willingness to sponsor me in this endeavor.

I wish to thank CAPT Steve Burris and CAPT Mark Converse for taking the time to sit and offer insight and experience into the world of effectively testing a weapon system.

I would like to thank the Anti-Radiation Missile Program at China Lake for constantly moving meetings out of the conference room to support my VTC classes.

Finally, I would like to thank my thesis advisors, Professors John Dillard and Walter Owen and Mr. Scott Weed. Their patience, insight, and words of encouragement have been key to the successful completion of this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. BACKGROUND

While the US has some of the most superior and highest quality weapons, repeated cost and schedule overruns routinely mar the product development timeline. These cost and schedule overruns normally lead to the destabilization of the programs and can cause a reduction in the unit buy (e.g., F/A-22), a reduction in funding to the program, or if severe enough, program cancellation (e.g., A-12 Avenger Program). Program Managers (PM) strive to prevent or minimize the situations that create this program instability. The Department of Defense (DoD) has funded and been the centerpiece of many studies to determine practices that optimize the product development timeline.

DoD has made concerted efforts to improve product development, but the Director of Operational Test and Evaluation (DOT&E) surmised in the FY01 and FY02 reports that the current trends in testing have systems beginning Independent Operational Test and Evaluation (IOT&E) while still in an immature status. “The cost of testing complex systems, as well as the risk of performance shortfalls delaying programs further, is motivating managers to skimp on testing.” (DOT&E, 2003, p.ii) As a result, a majority of systems in Operational Evaluation (OPEVAL) experience stops in testing or operational assessment failures. This was the case for the last High-speed Anti-Radiation Missile (HARM) software upgrade program. The system was sent back to the Developmental Test (DT) community after operational test failures. Upon returning to the Operational Test (OT) phase, it still received a failing grade for specific capabilities. This program failure, and those in other programs, translates into a combat capability lost or delayed resulting in increased risk to the Warfighter.

## B. PURPOSE

This research supports the strategic development of a Test and Evaluation (T&E) strategy for a weapon system through the analysis of common problems observed in a large-scale T&E program. This research allows the test community to proactively meet those challenges thus mitigating the risk of failure during OPEVAL. Historical facts show that a developing program risks failing OT if the challenges facing a test team during the developmental planning, execution, and analysis phases are not properly identified, assessed, and engaged. With reduced resources in today's acquisition world, it is unacceptable to ignore those challenges and fail to apply lessons learned from past programs. Throughout this paper, there are discussions and examples provided from various programs. The AGM-88E Advanced Anti-Radiation Guided Missile (AARGM) program is the primary case study.

The US Navy has recently entered into System Development and Demonstration (SD&D) for the acquisition of an upgrade to the AGM-88 HARM weapon system. This weapon system, designed to support the Suppression of Enemy Air Defenses (SEAD), is going through a major hardware and software upgrade. The new system is called the AGM-88E AARGM. This weapon system incorporates a new guidance and control section increasing its lethality and battlefield geographic specificity. In addition, it includes "net-centric" enabling capabilities by incorporating enhanced targeting and Weapon's Impact Assessment (WIA) information using the national support architecture. Incorporation of new weapon system components, program and test organizations, along with the use of a new contractor, means the T&E IPT must develop a test strategy that effectively and efficiently uses program resources to test the system at a level in DT that mitigates risk of failure during OPEVAL.

## C. RESEARCH QUESTIONS

This research provides insight to observed risks in OT. It then discusses strategies to mitigate some of those key risks. Research questions considered:

- What are some of the dominant factors affecting DoD testing?
- What have past studies offered as a means to help reform the T&E process?

- Can the application of commercial T&E Best Business Practices have a positive influence on government test process?
- What failure trends can be identified throughout test programs?
- What practices, processes, and planning can Developmental Test & Evaluation (DT&E) use to help acquisition programs succeed the first time they go into IOT&E?

## D. POTENTIAL BENEFIT FROM THIS STUDY

This study will identify varying facets of the T&E process and will be utilized by the AGM–88E program manager and the test team during their strategic development efforts. Understanding past T&E studies, best practices, and T&E lessons learned offers a wealth of knowledge to support the T&E strategy planning process. Currently there is a T&E strategy in the program's Single Acquisition Management Plan (SAMP). It does not offer the depth required to effectively develop a plan for testing. Additionally the current Test and Evaluation Master Plan (TEMP) is incomplete. While the document clearly defines the various stages of test and the time of execution, it is limited in depth with respect to a variety of essential test considerations. Some of these considerations include the conduct of test, firing scenarios, decision process, and asset allocation. Though it is not the scope of this research to directly answer all these considerations, this thesis will offer insight, allowing educated decisions to support the continued TEMP process. Built upon a solid foundation of lessons learned, the AGM-88E test program, specifically the DT program, will deliver a mature product to the operational test community. If that occurs, the program will meet its performance objectives, translating into an increased warfighting capability delivered on time.

In addition to the direct benefit that this thesis will provide the AGM-88E program, this study provides a source of documentation to support the test planning process for other acquisition programs. Although each program faces unique test challenges, there are common issues such as resource allocation that must be resolved. This thesis helps provide awareness to those generic issues, thereby increasing the knowledge base to effectively meet those challenges and limit their recurrence.

## E. SCOPE

The research focuses on identifying various elements to consider in the process of developing a test strategy to reduce risk during OT. The research is in five sections.

- The first section introduces the research topic and provides a brief discussion of the current issues affecting T&E and the general efforts over the years to develop a more effective set of practices supporting improvement to the product development timeline.
- The second section of the research focuses on the product development process and the relationship T&E plays in this process. It addresses the test approach and discusses the differences between commercial and DoD testing. The section additionally discusses the application of commercial best practices to the DoD T&E effort.
- The third section addresses lessons learned from previous programs. Trends are presented to the reader to highlight some key areas a tester must consider during the planning and execution of test. This section identifies the importance of proper resource allocations and requirements control. The section further discusses some of the key players with direct interests in a program and its success during testing.
- The fourth section introduces the AGM-88E weapon system case study. During this section, identified T&E risks are presented. Based on the research, a recommend path for the program is discussed.

- The final section presents recommendations, conclusions and offers suggestions for further study.

Based on the scope of this research, the reader will ascertain the general issues that affect the government team as it develops a product, with specific emphasis on the T&E process. Past efforts designed to foster a more effective and efficient DoD test process are introduced throughout the reading. Moreover, there is a discussion regarding the differences and difficulties applying commercial practices to DoD T&E. Furthermore, the research provides an opportunity to see that a conscious effort is being made early in a weapon system's development cycle to apply the best practices in test planning and execution of a major DoD acquisition program, to maximize the use and availability of limited resources. While sections of this thesis are specific to the AGM-88E system, they will provide enough generalities to be applied or at a minimum considered in the strategic test planning for other systems.

## F. METHODOLOGY

This thesis was developed using the following methodology:

- Literature reviews pertaining to T&E and product development;
- Interviews from representatives of various test agencies and former and current PMs;
- In-depth internet research pertaining to T&E, lessons learned, and acquisition documentation; and
- Lessons learned from personal practice and experience.

## G. TESTING A SYSTEM TODAY

Providing the best products to our military forces has always been a requirement in the US. By society's standards, it is unacceptable to send America's military into combat with systems that do not work as intended. Yet, despite this commitment, there is a consistent trend within the DoD acquisition community of not delivering products on time, within schedule, and within the proposed performance levels. There have been some great successes such as the Air Force's F-16 Fighting Falcon, but for every success there are the prominent failures. One notable failure was the Navy's A-12 Avenger program.

A past President's Scientific Advisory Committee stated the importance of T&E in the acquisition process. In the report the committee stated,

> We regard the creation of the testing and evaluation group as of the utmost importance, since we believe most of our previous failures to be prepared for wars we fight would have been thoroughly exposed had an adequate program of testing and evaluation existed. (Christie, 2002, speech)

The committee further identified the necessity of providing sufficient financial resources to the T&E organization to support adequate testing. They stated,

> The actual tests are very expensive and since the Testing and Evaluation budget in a Service is often in competition with funds for new equipment developments, we believe it is vital that the Test and Evaluation group in OSD have a substantial budget to allocate for tests. (Christie, 2002, speech)

Despite this recommendation, continued funding shortfalls prevent PM's from adequately executing a test program. Adding to the financial strain, the majority of funding to support the aging T&E infrastructure has transferred from institutional funding to program funding. This financial burden drives the PM to make compromises in efforts to test a developing system.

While support resources are being reduced, DoD continues to procure and drive for development and acquisition of more complex systems. These systems offer increased combat capability, but also increase the complexity of conducting T&E. Today's systems are no longer stand-alone systems to be developed and tested with a stovepipe mentality. The Navy's Sea Power 21 vision, which links information from various systems throughout the battlefield to support the Warfighter, as shown in Figure 1, has put a new interoperability requirement on all developing programs. This requirement and the basic system level requirements dictate a robust test effort that stresses available resources.

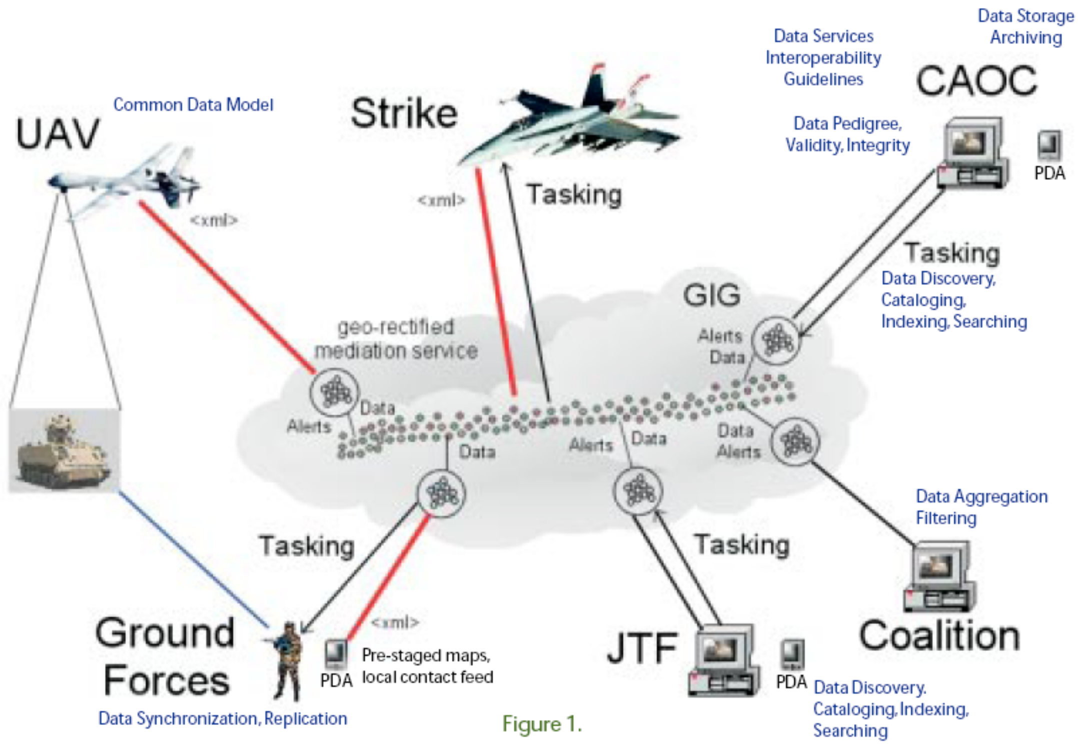


Figure 1. Net-Centric Warfare

DoD began many T&E process reforms to maximize the use of available resources and quickly transition programs from the design room to the war room. These reforms began in the early 1970s with a Blue Ribbon Defense Panel. This panel looked at acquisition policies and practices with respect to cost, schedule, and performance. Citing their findings, the current Director for OT&E (DOT&E), the Honorable Mr. Thomas Christie said, the panel concluded that the acquisition policy was, “highly inflexible . . . and also based on the false premise that technological difficulties can be foreseen prior to the detailed engineering effort on specific hardware.” (Christie, 2004, speech) The panel further recommended that prototyping, when applicable, be pursued in order to understand the technology and reduce the risk to the program’s development effort.

In 1986, the Packard Commission conducted another review of the acquisition process. In their final report, more than a dozen recommendations were proposed. Two recommendations focused on DoD’s T&E process. One recommendation supported the 1970 Blue Ribbon Defense Panel study. It emphasized, “proof of technology by building and testing hardware, including system prototypes where appropriate, and incremental development of subsystems and components.” (DOT&E, 2002, ¶12) The commission’s rationale, similar to the earlier study, was to reduce the technological risk of developing a new technology and to afford the PM an opportunity to conduct realistic cost estimates prior to full rate production. The final report further recommended the OT community play a larger role earlier in the development cycle and maintain that role throughout full-scale development. The Packard Commission believed that by exposing developers to the operational community earlier in the development cycle, there would be a higher probability the product performance would meet the operational needs.

In the 1990’s, with the Cold War over and financial resources for defense declining, there were increased reform efforts to reduce the time it took from program inception to operational use. Shorter product cycle times result in a reduction in costs and schedules while steadily improving combat capability. Since 1993, there have been seven initiatives affecting change in the acquisition process and supporting the efficient procurement of weapons systems. The final initiative was a complete cancellation and rewrite of the governing documents, which took place in 2002. (Rogers and Birmingham, 2004, pp. 47-48) Now in 2004, current regulations dictate defense acquisition plans give less guidance than ever before—less guidance on what to test; less on how to plan a T&E program; and less on how to document such planning in a TEMP. (Daly et al., 2003, p.1) Reduced guidance and more flexibility replaced the inflexibility sited in the 1970 study. The PMs and more specifically the test agencies now have an opportunity to more effectively plan and execute a test program, but if they are not properly prepared for such freedom of execution, they could unwillingly lead a program down an inefficient path.

## H. IMPACT OF PAST STUDIES

Since the Fitzhugh Commission in 1970, there have been numerous solutions proposed to improve the product development effort. As a result, “our program success rate has not greatly improved, lead time remains excessive, the drive for new untried technology still remains and delays new systems…cost overruns are still with us…and we ignore history." (Freeman, 1999, p.323) While there are isolated pockets of success stories, there continues to be a decline in the effectiveness of our test community to ensure a developing system makes it to the user on its first attempt through IOT&E. DOT&E Thomas Christie makes reference to the rush that PMs put on their test team to get the product out the door,

> We've rushed into operational testing when the results of DT&E have clearly shown us that we were not ready and that our chances of success were minimal. In essence, we have been rushing to failure. (Christie, 2002, speech)

The numbers support his claim. A report published by the Government Accounting Office (GAO) evaluated eight programs and their combined cost for development. In FY98, it was determined that it required $46.9 billion to complete the programs. In FY03, this estimate was adjusted to $71.6 billion, a cost growth of 53% in five years. (Levin, 2003) Financial resources within the defense budget cannot continue to support program cost growth of this magnitude.

## I. SUMMARY

If the focus remains on improving the performance of T&E since the 1970s, DoD's success rate is heading in the wrong direction. This research highlights some of the primary factors for this current situation. Aside from a continual change in operating procedures and guidelines, the T&E community has seen a dwindling budget, reduced support to maintain the range infrastructure, and reduced test expertise. All this occurs as the weapon systems and the technology become more complex, and the scenarios required for testing increase in complexity. Can major acquisition systems overcome these current challenges? To do so requires an understanding of the current issues that plague the community, an understanding of how the commercial world succeeds in their testing approach, and most importantly, an understanding of what can be learned from past programs, both the successes and failures.

# II. EVALUATING THE CURRENT CLIMATE

## A. INTRODUCTION

Various factors affect the ability for the tester to successfully perform his mission. The ability for the T&E community to recognize those challenges early in the planning process increases the probability of executing an effective test program. Recognizing common negative trends in T&E and understanding the results of past test improvement studies provides a solid foundation to build a program strategy. It is essential to understand the status of the country's range infrastructure in the support of T&E, as this will further provide a test planner awareness of potentially high-risk areas. Finally, it is important that a tester have knowledge about the historical relationship between the program office and the test community, as this will offer a glimpse into the impending difficulties in establishing an acceptable multi-organizational test program.

## B. COMMON TRENDS

While each Service has unique shortfalls associated with testing weapon systems, there are commonalities. DOT&E reported in their FY02 annual report that common areas, which resulted in T&E performance problems, include:

- Range encroachment;
- Failure to identify immature technology;
- Feedback loop breakdowns;
- Insufficient or inadequate developmental testing;
- Inadequate reliability testing;
- Poor software tracking and evaluation procedures;
- Insufficient prototypes and other test resources;
- Stability of engineering workforce;
- Inadequate evaluation of training;

- Hardware/software integration;
- Slow tempo of testing operations; and
- Insufficient interoperability of weapon systems.

(DOT&E, 2003, p.vi-ix)

Aside from problems such as range encroachment and the stability of the engineering workforce, the remainder can be controlled at the program level and mitigated with a well thought-out T&E strategy.

## C. T&E INFRASTRUCTURE

Range resources are essential for government programs to effectively test developing systems. These resources include facilities, airspace, land, targets, people, instrumentation, and data collection. Since the fall of the Soviet Union, funding shortfalls, divestiture, Base Realignment and Closure (BRAC), and a lapse in facilities maintenance have resulted in a decaying infrastructure unable to adequately support today's T&E demands. This shortfall has been identified by DOT&E as a major reason for inadequate testing of today's military systems. The Honorable Mr. Christie stated in a speech to the National Defense Industrial Association (NDIA),

> I am concerned that our T&E infrastructure is not in the best of shape needed to meet the challenges of the future. Failures of the acquisition process in the past, with all the program slips, have tended to ease the burden faced by the test ranges. Lord knows what would happen if all the programs that claimed to be ready for testing in 2004 actually showed up for testing. If the latest acquisition initiatives deliver what they hope for, then a greater fraction of programs should be ready for testing on or near their schedules. In this respect, I fear the T&E community might not be prepared for success in acquisition reform. (Christie, 2004, speech)

### 1. Range Resource

The Major Range and Test Facility Base (MRTFB) was established in 1974. Figure 2 provides a snapshot of the ranges that are part of MRTFB's coverage.

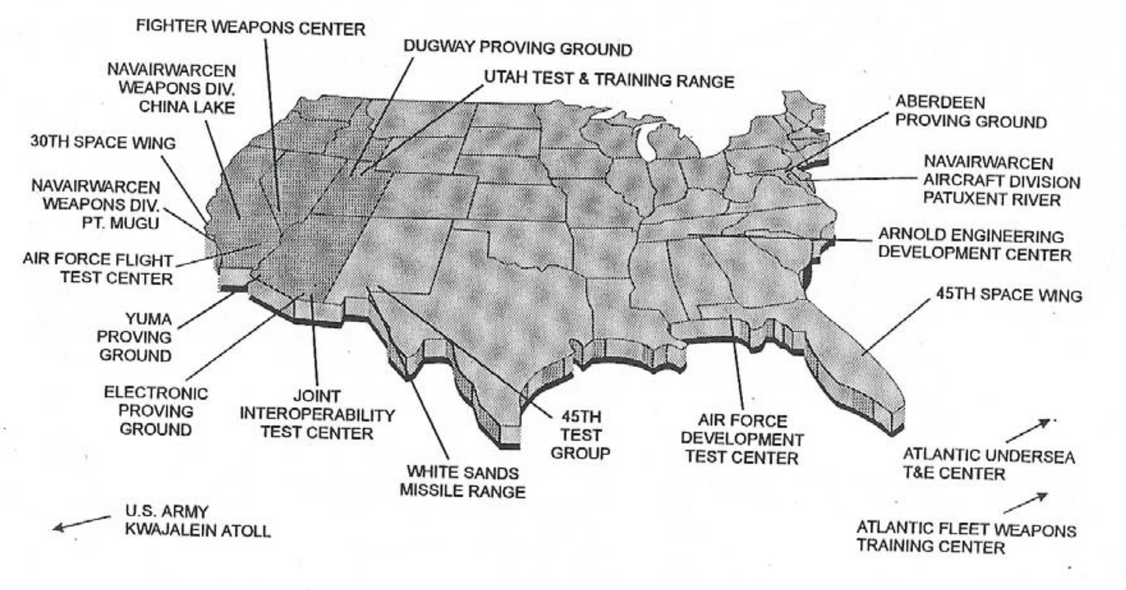


Figure 2. MRTFB Coverage

(Wascavage, 2004)

The MRTFB's governing policy states, "The MRTFB is a national asset that shall be sized, operated, and maintained primarily for DoD T&E support missions…"(DoDD 3200.11, 2003, p.2). Although this policy has not changed, MRTFB has seen its ability to support the policy become increasingly difficult. DOT&E points out that the primary reason for MRTFB's shortfalls are due to a lack of investment.

> Investment funding for the T&E infrastructure provided over the past 10 to 15 years has not kept pace with the identified T&E needs, severely restricting our ability to adequately evaluate new technologies such as stealth, command and control systems, hypersonic weapons, and missile defense systems. Funding for targets and threat simulator has also been sharply reduced. (DOT&E, 2002, p.II-4)

As a result of inadequate funding levels, the T&E recapitalization rate for the entire T&E infrastructure is 400 years. When evaluating the technical infrastructure the rate is 70 years. These rates are more than seven times that of the private sector. Secretary of Defense Rumsfeld set a goal to reduce the latter number to 35 years. (DOT&E, 2002, p.II-3) Failure to meet this goal will degrade the range support necessary to effectively test emerging weapon systems. The demands for higher levels of instrumentation and increased fidelity in target support are some examples of the shortfalls that affect programs because of the long recapitalization timeframe.

Reliability of the systems supported by the ranges is important to developing programs. As funds are reduced and systems begin to age, the cost to maintain and repair negatively affects the ability of the tester to complete the mission. Two examples of the impact to a program due to test infrastructure shortfalls include:

- A failure in a motor for a wind tunnel at the Arnold Engineering Development Center, TN resulted in a reduced capability for over seven months.
- Intermittent failures in radio frequency (RF) emitters at NAWS China Lake, CA have affected the ability of the ARM Weapons Office to conduct flight-testing in the development of software upgrades to the HARM weapon system.

These support shortfalls lead to test delays resulting in Fleet delivery delays or, if severe enough, the cancellation of the effort.

Range encroachment has also threatened the MRTFB. As US population and urban development continue to grow, DoD ranges continue to feel the effects. Ranges that once were isolated are now finding housing developments near range boundaries. Airspace for the military, whether for testing or training, is continuing to decline because of pressure from commercial industry. Over the course of the last 10 years the China Lake airspace has seen increased restrictions as a result of both urbanization and commercial air traffic. One example is the increased difficulty to conduct low altitude testing within the Sierra Mountain range. Noise complaints from the growing local population have resulted in flight restrictions keeping aircraft restricted either from designated areas or at altitudes that are not operationally representative. The unfortunate circumstance with this current trend is that emerging weapon systems are requiring more airspace to effectively test. This diametrically opposed flow results in programs not having the range space to fully test the system. A developing weapon system, sponsored by the Office of Naval Research (ONR) as a Future Naval Capability (FNC), which incorporates the use of ram jet technology, will be conducting live-fire testing in the coming years. The current concern is that the distance this weapon system can travel cannot be supported by any land range were data collection is best. The test team assigned to the program is currently addressing this issue for possible alternatives. The Missile Defense Agency (MDA) is also addressing range limitations.

> Addressing a shortfall identified by DOT&E, the MDA is minimizing flight test restrictions by adding more intercept regions and launch locations to add greater realism to its tests. MDA is expanding the test range infrastructure to add five intercept regions and target and interceptor launches out of new locations. (Government Accounting Office-04-254, 2004, p.3)

Congestion in the frequency spectrum due to the commercial sector's desire for increased frequency usage is also becoming a concern for the test community as it adversely affects military testing. DOT&E has documented that ranges already delay tests because they do not have enough frequency spectrum to run multiple tests simultaneously. These conflicts are common at the China Lake range facility and typically dictate which programs receive range periods. Tight scheduling of instrumentation frequencies can create problems for programs. When granted a frequency transmit time, it typically covers the prescribed range period. It does not normally cover the time spent on deck, which for some programs is a critical phase to determine if the aircraft should launch. If there is a restriction to transmit during ground operations, programs must either launch with an unknown system status or delay launch, thereby reducing the overall test time available for the range period.

## 2. People

Shortage of personnel has dramatically affected the T&E mission. Loss of government, military, and contractor personnel from the ranks of the T&E workforce has created holes in the support structure, reducing corporate knowledge, leadership, and dedicated blue-collar labor. A function of funding, the workforce levels have dropped drastically over the last 10 plus years, shown in Figure 3, while the workload has increased.

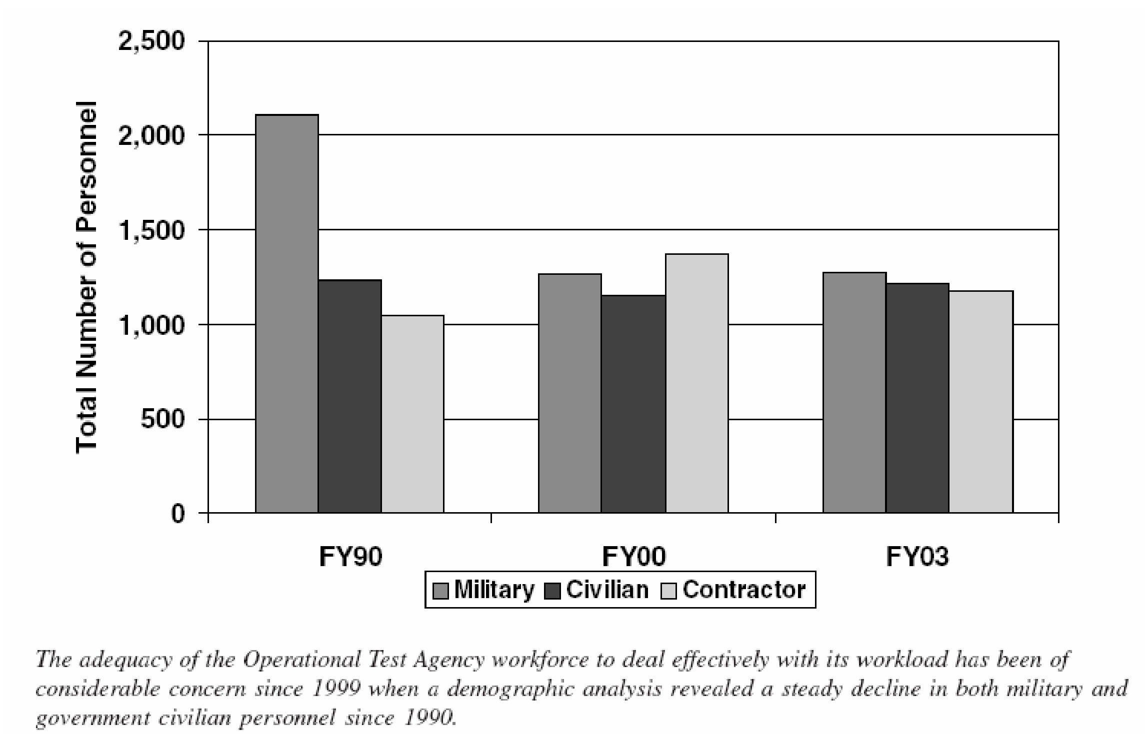


Figure 3. MRTFB Workforce Levels

(DOT&E, 2004, p.341)

The impact of government and military manpower shortages require that more contracting personnel be assigned to the program, increasing overall cost. This was the case in the F-22 program where a reduction in government workforce attributed to a plus-up in the Lockheed Martin contract.

Manpower shortages dramatically affect the ability of the military test community to actively participate in the development of a weapon system.

> There is urgent need to bring military personnel back into the infrastructure so that systems undergoing developmental test can have the benefit of direct soldier input. There is increased emphasis on providing earlier feedback to the development process; however, user participation is diminished. Military users and operators must be restored to developmental testing in order to enhance the effectiveness of test programs. (DOT&E, 2003, p.318)

These shortages in personnel, specifically military, are prevalent at the Navy's DT and OT squadrons. Presently at the VX-31 Developmental Squadron, officer manning is at 85%. (VX-31 Report, 2004, p.14) Within these commands, military testers find themselves handling multiple programs. This high workload results in a reduced performance level, thereby raising the probability that operationally related problems will be overlooked, increasing program risk.

### 3. Targets

"The current inventory of targets does not adequately replicate emerging threats. Adequate operational testing of new weapon systems requires targets possessing significantly greater threat fidelity." (DOT&E, 2003, p.328) DOT&E further went on to highlight the shortfalls in targets in their FY03 report. "Testing has been delayed or not completed due to the absence or unreliability of available aerial targets." (DOT&E, 2004, p.343) In that report, specific types of targets were identified as being unsuitable for future use. The first was the fidelity of DoD's aerial targets. Currently the QF-4 target aircraft, which is supported by VX-30 at NAS Point Mugu and is the primary target aircraft for Air-to-Air (A/A) testing, will soon be divested. This unique capability does not have a replacement. The other issue with the QF-4 is the type of target it represents. A Vietnam era aircraft, it does not adequately represent the aerial threats that our Warfighters face today or in the future. Another key target asset that has not been replaced is the Self Defense Test Ship (SDTS). This ship is integral in the development of systems that include Ship Self Defense Mark 2, Rolling Airframe Missile, Evolved Sea Sparrow Missile, DD(X), and CVN 21.

Operationally representative targets are essential to verify system performance during test. Shortages or lack of availability will limit the knowledge gained by the test team for test events. To ensure targets are representative, the test community must adhere to the accreditation process identified by DOT&E. Since the cancellation of DoD 5000.2-R, which stated, “representative threats must be validated by DIA or the DoD Component Intelligence Agency, and approved by DOT&E.” (DoD 5000.2-R, 2002, p.58) Programs must maintain early communication with DOT&E to establish the acceptable Verification, Validation and Accreditation (VV&A) process to avoid surprises late in testing.

#### 4. Instrumentation and Data Collection

> The ongoing military transformation requires the T&E community to be prepared to test more sophisticated systems employing more advanced technology. Without the resources and funding required to sustain, maintain, and modernize T&E, we face the inescapable conclusion that T&E will reach a point in the foreseeable future where the quality of testing and the information provided will deteriorate below reasonable and acceptable limits. (Gehrig et al., 2002, p.58)

The Joint Strike Fighter (JSF) fly-off may have already highlighted the level of degradation and availability of instrumentation and data collection service presently available. During that time, the two competitors approached the use of the range and its ability to provide instrumentation data in two very different fashions. One contractor followed the standard approach by relying on the already established data collection bays within range control. This approach led to scheduling delays or lost test events due to range support availability conflicts. The other contractor developed a unique data collection van. This remote and mobile facility allowed this contractor to more effectively achieve and complete test events by reducing the reliance on the range bays and personnel needed to operate them. It afforded this contractor a broader range of test times that would have otherwise not been available due to other test program conflicts or range personnel work schedules. (CAPT Burris, 2004, interview)

### 5. DoD Test and Resource Management Center (DTRMC)

The DTRMC is a recently established organization, which reports directly to the Under Secretary for Acquisition, Technology, and Logistics (USD(AT&L)). The DTRMC stems from a recommendation made by the Defense Science Board (DSB) task force in 1999. "OSD and the Services should work together to develop a plan whereby T&E resource management is strengthened and brought under coherent control." (Defense Science Board, 1999, p.23) Their function is to develop and maintain a strategic plan for T&E and to certify the adequacy of T&E resources. DOT&E believes that the establishment of such an organization is essential to support T&E as new and innovative programs begin to enter the acquisition pipeline. They add that the DTRMC will focus scarce T&E investment resources toward the most critical needs and address future needs. (DOT&E, 2004, p. 337)

## D. DOD TEST PHILOSOPHY

### 1. Test Communities

Within the DoD acquisition community, there are multiple test agencies and commands each with different responsibilities. While numerous, they are in place to support the two primary test communities: DT and OT. Agencies/commands focused on DT product development determine whether a system meets the technical specifications as defined in the contract and system specification. These specifications are the basis of the Critical Test Parameters defined in the TEMP.

The responsibilities for the DT community as stated in DODI 5000.2 dated May 12, 2003 are:

- Identify the technical capabilities and limitations of the alternative concepts and design options under consideration;

- Identify and describe design technical risks;
- Stress the system under test to at least the limits of the Operational Mode Summary/Mission Profile, and, for some systems, beyond the normal operating limits to ensure the robustness of the design;
- Assess technical progress and maturity against critical technical parameters, to include interoperability, documented in the TEMP;
- Assess the safety of the system/item to ensure safety during OT and other troop-supported testing and to support success in meeting design safety criteria;
- Provide data and analytic support to the decision process to certify the system ready for IOT&E;
- Conduct information assurance testing on any system that collects, stores, transmits, or processes unclassified or classified information.
- In the case of IT systems support the DoD Information Technology Security Certification and Accreditation Process and Joint Interoperability Certification process;
- In the case of financial management, enterprise resource planning, and mixed financial management systems, the developer shall conduct an independent assessment of compliance factors established by the Office of the USD; and
- Prior to full-rate production, demonstrate the maturity of the production process through Production Qualification Testing of LRIP assets.

(DoDI 5000.2, 2003, pp.26-27)

On the other side of the test spectrum is the OT community. This community evaluates the effectiveness and suitability of a system in a realistic operational environment. The objectives of the OT&E phase as defined in the DoDI 5000.2 are:

- OT&E shall determine the operational effectiveness and suitability of a system under realistic operational conditions, including combat; determine if thresholds in the approved Capabilities Production Document (CPD) and critical operational issues have been satisfied; and assess impacts to combat operations;
- Typical users shall operate and maintain the system or item under conditions simulating combat stress and peacetime conditions;

- The independent Operational Test Agency (OTA) shall use production or production representative articles for the dedicated phase of IOT&E that supports the full-rate production decision (or for Acquisition Category (ACAT) IA or other acquisition programs, the full-deployment decision);
- Hardware and software alterations that materially change system performance, including system upgrades and changes to correct deficiencies, shall undergo OT&E;
- OTAs shall conduct an independent, dedicated phase of IOT&E before full-rate production to evaluate operational effectiveness and suitability, as required by reference; and
- All weapon, Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR), and information programs that are dependent on external information sources, or that provide information to other DoD systems, shall be tested and evaluated for information assurance.

(DoDI 5000.2, 2003, pp.27-28)

The Key Performance Parameter (KPP) as defined in the Capabilities Description Document (CDD), formerly known as the Operational Requirements Document (ORD), is the metric that the OT community uses. All systems under test must meet the KPPs in order to establish a foundation for operational test success.

Historically the two testing phases were conducted in very structured and separate stages of a test program. DT executed their functions and then when complete, transferred the program to the OT community. With acquisition reform, there have been attempts to integrate the two phases of test. This integration would reduce the test repetition between organizations, identify deficiencies earlier in development, and clearly identify OT recourses, thereby reducing the overall product development timeline. This approach supports one of the recommendations by the DSB.

> Each of the Service DT & OT organizations should be consolidated; to include integrated planning use of models, simulation, and data reduction. Planning should be totally integrated, and the OSD T&E organizations consolidated. There should be integrated use of

> models, simulation, and data reduction. Except for limited dedicated OT&E, contractor and government testing should also be integrated. (DSB, 1999, p.23)

### 2. Test Approach

PMs must face many challenges throughout their tenure. They come to the position with a variety of programs at various stages of development, and they must ensure that they can preserve each one. They are faced with this formidable challenge due to the foundation that a majority of DoD programs are built upon. They see exaggerated optimism in scheduling and unrealistic estimates in budget planning.

> Both industry and DoD program managers have suffered from a contagious trend of unmerited optimism in defining and supporting both cost and schedule program risks, especially across the most complex programs such as V-22, F-22, and Comanche. The initial program baselines were built around making the programs fit inside a constricting cost and schedule box vs. designing program plans within flexible boxes to accommodate the many unknowns associated with complex integration initiatives. (Birmingham and Rogers, 2004, p.55)

As a result of this unstable foundation, PMs normally delay system testing until late in the development cycle. This affords the program time to let technology catch up to the requirements and prevents unwanted attention from decision-makers who may be interested in diverting funds. This approach, while short sighted and extremely risky, is the path that the current acquisition process forces a PM to follow. The GAO noted this approach during a study of a major aircraft development program.

> Our work has shown that numerous weapon system programs suffer from persistent problems associated with late or incomplete testing. This practice pushes the burden of discovery late in development when problems become very costly to resolve. We also found that testing operated under a penalty environment that creates perverse incentives. For example, if tests were not passed, the program might look less attractive and be vulnerable to funding cuts. Managers thus had incentives to postpone difficult tests and limit open communication about test results. These represent

widespread and systemic problems within the Department that must be addressed. (GAO-01-369R, 2001, p.2)

Another key element that drives the PM to this avoidance test strategy is DoD's failure to properly recognize immature technology when an acquisition program begins. There is confidence that given the right amount of time the technology will be there when needed in a program's development cycle. This eventually results in program cost overruns and schedule delays. GAO reported,

> The competition for funding at the time of launch encourages aspiring DoD programs to include performance features and design characteristics that rely on immature technologies. Untempered by knowledge to the contrary, the risks associated with these technologies are deemed acceptable. Because production can be 15 years from the launch decision, it is difficult for production realities and concerns to exert as much influence on a DoD product development as they do on commercial products. Instead, design features and performance are more dominant. More unknowns are accepted on a DoD program, and their attendant risks are often understated. This combination, which can be devastating to a commercial business case, can help a weapon system program get launched and survive. (GAO-98-123, 1998, p.15)

| | | Product development | |
|---|---|---|---|
| **Product development and associated technologies** | **TRL at program launch** | **Cost growth** | **Schedule slippage** |
| Comanche helicopter | | 101 percent[a] | 120 percent[a] |
| Engine | 5 | | |
| Rotor | 5 | | |
| Forward looking infrared | 3 | | |
| Helmet mounted display | 3 | | |
| Integrated avionics | 3 | | |
| BAT | | 88 percent | 62 percent |
| Acoustic sensor | 2 | | |
| Infrared seeker | 3 | | |
| Warhead | 3 | | |
| Inertial measurement unit | 3 | | |
| Data processors | 3 | | |
| Hughes HS-702 satellite | | None | None |
| Solar cell array | 6 | | |
| Ford Jaguar | | None | None |
| Adaptive cruise control | 8 | | |
| Voice activated controls | 8 | | |

Figure 4. Cost and Schedule Experiences on Product Development (GAO-99-162, 1999, p.27)

Figure 4 reflects the impact that immature technology at program inception can have on a program cost and schedule. Technology Readiness Levels (TRL) definitions are provided in appendix A. TRL numbers represent product maturity levels for key technologies. While there is no requirement to accept a desired number as a benchmark for inclusion in a program, GAO studies have identified that DoD typically accepts readiness levels below that of commercial firms, as shown in Figure 5, resulting in higher cost and schedule overruns. Ultimately DoD cancelled the Comanche helicopter program in the spring of 2004.

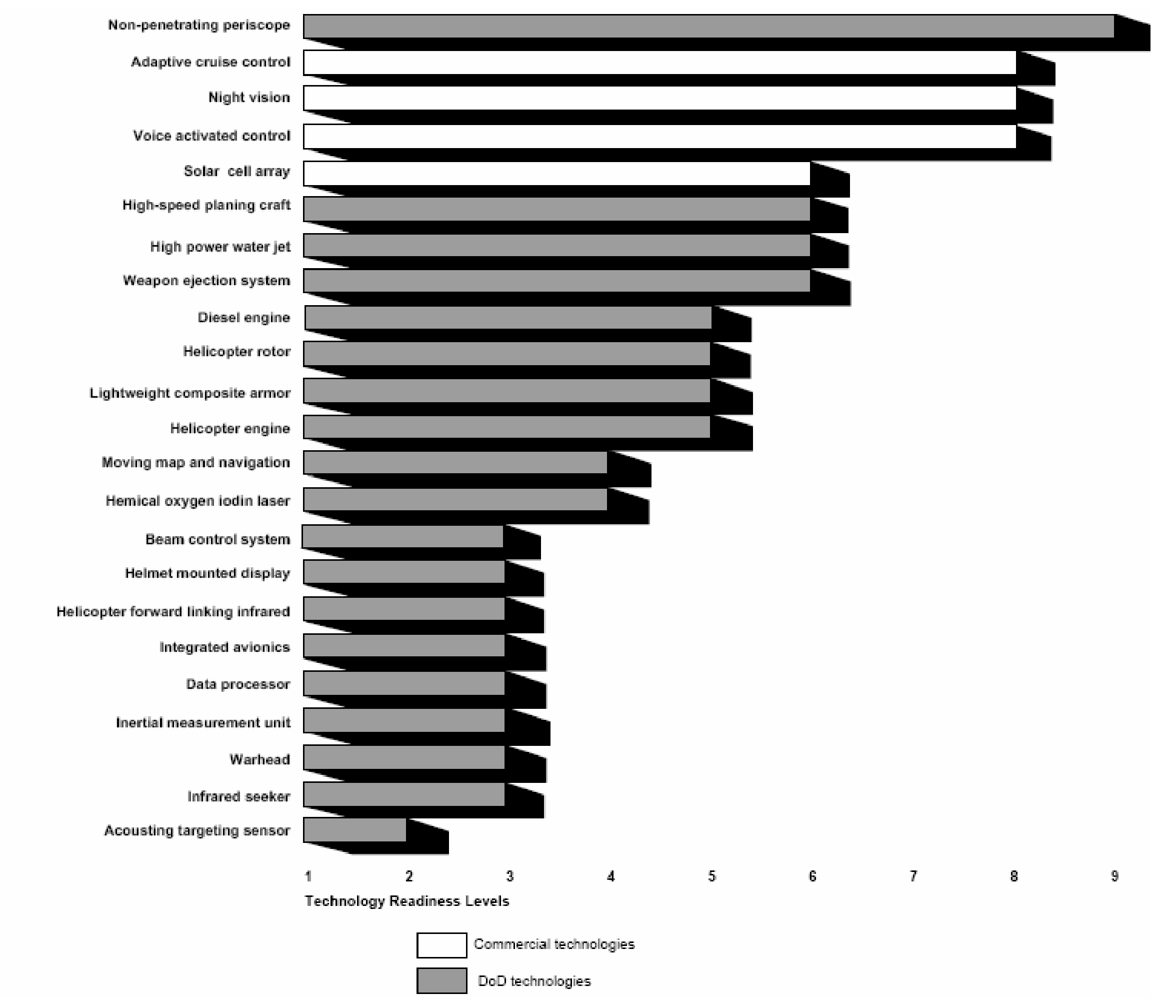


Figure 5. Readiness Levels of Technology at Time of Program Inclusion (GAO-99-162, 1999, p.26)

Another DoD development program that suffered from immature technology during the development phase was the A-12 Avenger. The immature technology could not support the proposed design or requirements. The amount of composites that were necessary to support the overall structure for a carrier environment while maintaining the stealth capability resulted in weights that exceeded the specification by almost 30%. The composite technology was so immature that General Dynamics and McDonnell Douglas would have had to develop it during full-scale development, because they had limited experience in building large structures using composites. (Pike, n.d., ¶ 7)

Faced with schedule slippages due to technology delays and the increased budget that accompanies these slips, PMs are prone to hold off testing until late in the development cycle. During this non-test time, PMs use Power Point presentations and engineering studies to show progress and begin developing a case that when testing does begin, the test phase should be expedient with no major systems failures. Often this is not the result. This approach to testing creates late cycle development problems that could have been identified and corrected at lower cost earlier in a program's schedule had the proper subsystem testing been performed. Theater High Altitude Area Defense (THAAD) Program used this conservative approach. "Instead of break it big early philosophy, program officials waited until flight testing to stress components and subsystems. As a result, key subsystems were not sufficiently matured for integration and flight testing." (GAO-00-199, 2000, p. 34)

Failure early will create a perception of program trouble. This can allow other programs and adversaries to lobby for the cancellation or reduction of funds for the respective program. Since program funding is typically unstable and consistently up for reviews, PMs attempt to postpone any chance for perceived failure for as long as possible by delaying or canceling test events. Missile Defense Program follows this philosophy. GAO noted that, "MDA is generally not addressing DOT&E's proposal for ground testing…MDA deferred testing at the facility to fund other priorities." (GAO-04-254, 2004, p. 3) Senator Jack Reed (D-RI) further supported GAO,

> This report confirms that rather than thoroughly testing the missile defense system, the Administration is blindly spending billions of dollars every year with the exclusive goal of deploying system by September even if that system is ineffective and it capabilities untested. (Reed, April 23, 2004)

DoD testing methodology does not offer a true understanding of product maturity. The methodology supports a pass/fail system. If the test event equates to a pass, then the program survives. Conversely, if there is failure, interest within DoD and potentially Congress increases. DOT&E recommended that the test community shift away from this black and white metric and evaluate based on knowledge gained. The Honorable Mr. Christie reinforced this concept by stating,

> Testing is for learning! That may sound somewhat trite, but how often have we strayed from that dictum and reflected the proverbial Pass/Fail mentality we're so often accused of. (Christie, 2002, speech)

Because of the pass/fail philosophy, complete integrated system tests are normally held off until late in the program's developmental stages or the complexity of test scenarios are limited to ensure a successful test. To Cite the GAO report on the ballistic missile defense, "no component of the system to be fielded by September 2004 has been flight tested in it deployable configuration." (GAO-04-254, 2004, p.4)

The overall outcome of such strategies leads to the identification of major system problems late in the development cycle. This delays knowledge about the program's product maturity. A GAO report that assesses major DoD programs states,

> The difference between highly successful product developments—those that deliver superior products within cost and schedule projections—and problematic product developments is how this knowledge is built and how early in the development cycle each knowledge point is attained. (GAO-03-476, 2003, p.4)

PMs must also face another burden. DoD has reduced institutional funding to support the test ranges, as shown in Figure 6, and placed the responsibility upon the PM. Institutional funding now accounts for less than 40% of the financial resources that go to support the range infrastructure.

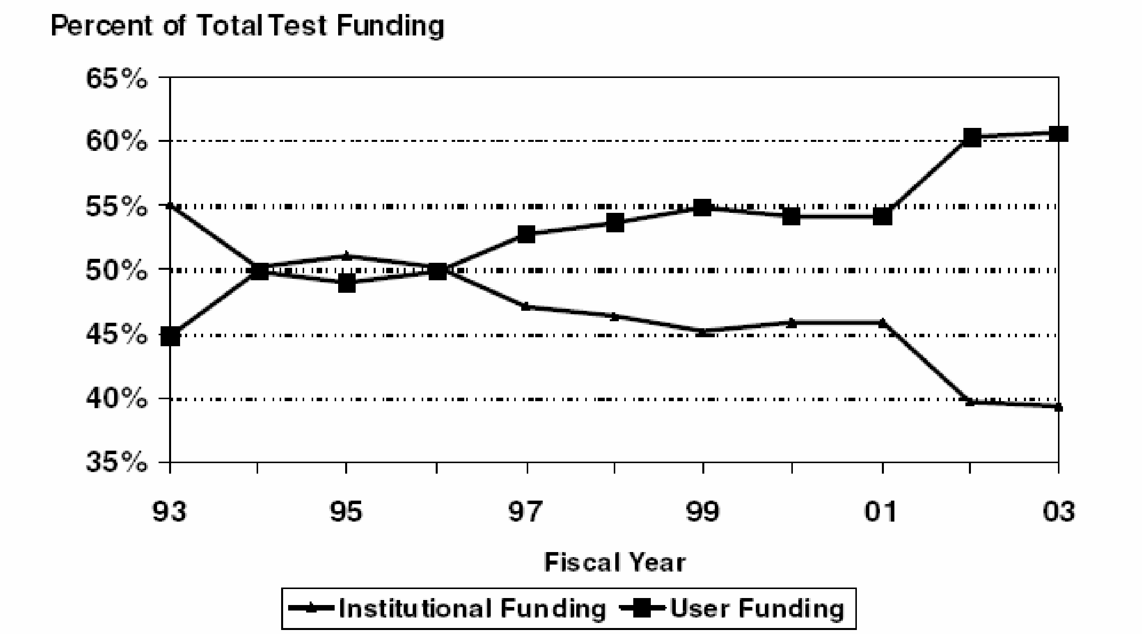


Figure 6. MRTFB Funding Responsibility
(DOT&E, 2004, p.340)

The transfer of financial responsibility contributes to the reduction in the DT effort. The Honorable Mr. Michael Wynne during testimony to the US Senate Committee on Armed Services identified this as a concern. “We are concerned with the continuing problem surrounding overhead costs and their impact to program managers when they use the test ranges and facilities.” (Wynne, 2002, p. 5) Congress directed DTRMC to evaluate and develop a strategy reversing this trend.

### 3. Test Culture

As testers report poor test results to the PM, there is a natural tendency for the PM to become defensive. The PM views these failures as an impact to his or her schedule and budget. As a result, the inherent nature of the PM-Tester relationship, even before testing begins, is adversarial. “Too often testing is seen as the spoilsport, the bearer of bad news, or at least cold reality – and facts and figures that aren’t as glowing as the program manager would have wished.” (Johnson, 2001, p.69) Other studies site this negative relationship.

> Research has found that a negative test culture exists in many PMOs, and this culture may have been the basis of testing problems. Several PMOs, and sometimes contractors, have displayed a negative attitude toward testing, testers, and analysts. The representative causes noted for this problem included the acquisition process itself, lack of PMO understanding of test and analysis capabilities and constraints, and the assumption that testers and analysts always require more or excessive testing. However, it was also found that some testers and analysts have earned poor reputations among program offices by conducting tests that appeared to add no value to the process or testing for weapon capabilities that were beyond the design requirements. (Hoivik, 2000, p. 35)

In order to promote positive test culture, there needs to be an honest and continuous flow of communication between the test agencies and the PM.

### 4. DoD Summary

The DoD test structure delineates between the DT and OT communities. Within each division, there are multiple test agencies and commands. The current direction by the acquisition community is to integrate these communities to promote an overall reduction in product development time. This should then translate in cost and schedule savings with reduced potential for program failure during the independent operational assessment. Even if DoD is successful in integrating the two communities during the test effort, they still face challenges to overcome in the basic test philosophy. DoD continues to accept low technology readiness for key technologies at program inception. They further delay testing until late in product development, thereby reducing system development knowledge and increasing overall risk. This approach is contrary to the desires of the test community, specifically the DT community, and as a result fosters a negative test culture.

## E. COMMERCIAL TEST PHILOSOPHY

The commercial industry has a much different approach with respect to product testing. They view testing as a learning opportunity and a means to evaluate progress in the transformation of a vision into a product. They test vigorously early in development to avoid late-cycle development problems. There are three distinct learning phases in commercial testing: (1) components work individually; (2) components work together as a system in a controlled setting; and (3) components work together as a system in realistic settings. (GAO-00-199, 2000, p.5) At the basic level, these cycles are not much different from the cycles in DoD testing. What is different is the approach. Distinct differences from DoD are the level of technical maturity required at program initiation and the test approach. Commercial firms are adverse to include technology that is not mature. Immature technology creates unnecessary risk and can result in schedule delays or cost overruns. Boeing followed this approach during the development of the 767.

> Boeing’s conservative approach was illustrated in the 1970s and 1980s when it decided not to include in its 767 more advanced systems such as fly-by-wire, fly-by-light, flat panel video displays, and advanced propulsion systems. Even though the technology existed, Boeing did not believe it was mature enough for the 767. (Battershell, 1995, p.215)

Testing is also approached differently. When tests are structured and executed, it is to gain knowledge and to help improve the product. Firms consider a test a failure if there is no increase in product maturity knowledge. Program managers for the 777-200 aircraft, a highly successful development and test effort, considered problems resulting from test as “gems to be mined” (GAO-00-199, 2000, p.8) and believed earlier identification resulted in less expensive fixes.

Commercial firms have a stake in delivering products on time and within defined performance standards. Their motivation is profit. If product’s performance, schedule, or cost metrics are below expectations, the impact results in a reduced market share, disgruntled customers, and lower profits for the product. This reality provides motivation to industry leaders such as Boeing, Intel and Dupont. Each company has experienced the ill effects of a poor T&E culture that resulted in the discovery of product problems late in the development cycle. In some cases, discovery occurred after product delivery.

### 1. Boeing’s Lesson Learned

Boeing’s educational awakening to the value of testing early and aggressively was a result of problems experienced in the development of their 747-400 models. Significant problems identified late in the aircraft’s development cycle, due to design and requirements changes, resulted in ineffective testing, late deliveries, and eventual service problems.

> Typically, engineers were still designing when manufacturing began, and they kept making changes as problems subsequently came to light on the factory floor, on the flight line, and even in the customer’s hands after the plane was delivered. For example, when Boeing delivered the 747-400 to United in 1990, it had to assign 300 engineers to get rid of bugs that it hadn’t spotted earlier. United was not happy with Boeing’s late delivery of the 747, nor with the additional costs the airlines sustained in rescheduling flights and compensating unhappy customers as a result of maintenance delays. Boeing was deeply embarrassed by delivery delays and initial service problems for its 747. (Battershell, 1995, p.217)

As a result, a new development and test approach drove management during the development of the 777-200 aircraft. This approach fostered an increase in the scope of testing thereby identifying problems early in the cycle. The result was a product delivered on time within performance requirements and with a 60% reduction in changes, errors, and rework as compared to their previous aircraft programs. (GAO-00-199, 2000, p.23)

### 2. Intel's Lesson Learned

Intel's experience stemmed from not properly analyzing and learning from test data. With the development of a microprocessor, Intel conducted a test that indicated a problem with higher-level mathematical functions. Based on the test data, Intel concluded that the impact would be minimal to the consumer as the occurrence of the failure would be rare. This analysis led the company to release the microprocessor with a known flaw. Unfortunately, a miscalculation with respect to the rarity of the failure resulted in the company having to replace more than a million microprocessors at a cost of $500 million. (GAO-00-199, 2000, p.25) Intel corrected their flawed T&E and analysis approach and today is successful with Pentium processor development. Intel reported that the reason for the "bug" being allowed to hit the market was a result of testing concluding too early. They determined that if the development team had exercised the system longer, the effects of the computer bug would have been identified before market release. Similar to Boeing, Intel increased the amount of validation testing conducted to identify problems, and with the increase in effort, they increased the amount of personnel support. The latter is typically difficult to do in a DoD program due to funding constraints and the availability of qualified personnel. Intel's increased focus, with respect to T&E, for its microprocessors has resulted in an increased product release rate.

### 3. Dupont's Lesson Learned

Dupont's realization about its poor T&E approach was a result of an internal analysis regarding its product development effort. It identified that it was taking twice as long to deliver a product to the customer as its competitor. This resulted in a loss of millions of dollars in revenue. It determined that the company's philosophy on test failures was driving them to identify problems late in a product's development life cycle. This led to late corrective actions at high costs, a process familiar to DoD. As a result, the company changed their paradigm that test failures meant a bad product. They now have adopted the approach that test failures are a means to learn more about the product's development. Their philosophy is "if a test does not lend any new information about the system's maturity then it is considered a failure." (GAO-00-199, 2000, p. 45) With resources and costs of testing rising, this approach ensures effective utilization of limited test resources.

#### 4. Management and Tester Relationship

Companies further view the testers as equals in the product development effort. Their input is valuable to the successful development of a product. This positive relationship helps foster test team motivation, since they feel they are members of an organization trying to make the product succeed. This relationship further provides a personal boost to each member as it helps instill the concept that his or her input is important to the development of a product. As highlighted in a previous section, this working environment is not always present in DoD efforts. PMs sometimes view testers as roadblocks, and testers sometimes create difficulties for the PMs by not properly testing or evaluating a system due to their ignorance system requirements.

#### 5. Commercial Summary

The overall commercial philosophy on testing is quite different from the DoD approach. Commercial firms have a knowledge-based testing approach. They effectively and efficiently attempt to use the testing resources available to provide knowledge about the program's maturity during the development effort. They focus on testing systems hard early in development in hopes of identifying trouble areas. Early identification of deficiencies will allow fixes at a reduced cost. The commercial sector is able to take this approach because of the means by which they fund a program. Unlike the complete financial support normally given by a commercial firm, the government approach requires that programs continuously defend their budget. This results in PMs taking a less aggressive test approach early, thereby delaying the identification of problems until late in the system's development effort. During this time of late discovery, resources are normally low, and more are required to correct the identified flaws. While this is the current practice, DoD is driving PMs to be more aggressive in the knowledge-based testing environment. Embedded in the DoD 5000.2, there is guidance that supports the philosophy discussed in the commercial industry.

> Knowledge-Based Acquisition. PMs shall provide knowledge about key aspects of a system at key points in the acquisition process. PMs shall reduce technology risk, demonstrate technologies in a relevant environment, and identify technology alternatives, prior to program initiation. They shall reduce integration risk and demonstrate product design prior to the design readiness review. They shall reduce manufacturing risk and demonstrate producibility prior to full-rate production. (DODD 5000.2, 2003, p.5)

## F. DOD STUDIES

Two major studies, sponsored by DoD, were undertaken in the late 1990s and in the early part of the new century. The first, developed by the DSB Task Force on Test and Evaluation was chartered by the Under Secretary of Defense (Acquisition and Technology) (USD (A&T)) in 1998. The DSB was tasked to review all activities relating to T&E within DoD. This monumental task culminated in a final report on T&E in 1999. The focus of the report was to identify the current state of T&E and offer recommendations to overcome any identified shortfalls. The concern that drove this report was the expectation that procurement of major programs would be on a steady increase. This trend would put a strain on the current range infrastructure and the overall RDT&E budget, necessitating a push to become more efficient in the business of T&E. The second report, directed by the Deputy Director, DT&E USD (AT&L), focused on industry best practices and their applicability to DoD DT&E.

### 1. Defense Science Board Study

The DSB's directives that guided the study were:

- Examine new and innovative ways that the T&E community can better support its users;
- Find new ways to integrate operational testing into the overall system development process;
- Consider the special problems associated with T&E of the "systems of system" which increasingly comprise critical parts of our military capability;
- Identify and quantify the current and future needs of the Department's T&E capabilities and resources; and
- Recommend specific and quantified changes.

(DSB, 1999, p. 10)

Their research and analysis offered observations and recommendations to improve the T&E process. The findings include:

- The focus of T&E should be on how to best support the acquisition process;
- T&E planning with operational test personnel should start early in the acquisition cycle;
- Distrust remains between the Program Management and test communities;
- Contractor Testing, Developmental Testing, and Operational Testing have some overlapping functions;
- Independence of evaluation of test data is the essential element, not the taking of the data itself; and
- Response to perceived test "failures" is often inappropriate and counter productive.

(DSB, 1999, pp. 1,2)

**Recommendation Review**

The focus of T&E should be on how to best support the acquisition process. The test community must establish a test approach that supports learning and confirming the systems performance at various stages of product development. As indicated in a previous section, DoD has encouraged this approach in the recent DoD 5000.2 with Knowledge Based Acquisition.

T&E planning with operational test personnel should start early in the acquisition cycle. First, by interfacing with the operational testers and users earlier in the test process, the test team can confirm that they understand the requirements. This will enable them to design test scenarios that evaluate the system based on expected Fleet/Field usage. This has been a recommended approach by DOT&E. While operational scenario testing is expected from the OT community, DOT&E proposes that this philosophy flow into the DT paradigm. "We must reinforce the principle that systems that go to war must be tested the way they will be employed." (DOT&E, 2003, p. iii) Second, testers, specifically the OT community, must attempt to participate early despite resource constraints. They must also not consider that early involvement will result in losing their independence in test. This early involvement was a major element for the success experienced with the F/A-18E/F test program.

> Active participation of VX-9 (Navy OT squadron) in the Integrated Test Team ensured that operational insights were always readily available to the developing organizations. The benefits of this close coupling were demonstrated as the program discovered and then overcame a flight problem referred to as wing drop. As modifications were installed to counter this phenomenon, the active participation of operational pilots provided rapid feedback as to whether the phenomenon interfered with mission conduct. This synergy between operational insight and developmental effort allowed alternative designs to be quickly evaluated. A production fix was determined, and a potentially major deficiency was rapidly corrected. (Institute for Defense Analysis, 1999, p.11)

Distrust remains between the Program Management and test communities. This has been a recurring discovery from a variety of sources highlighted in this research. Similar to the commercial sector philosophy, PMs should view the test community, both DT and OT, as members of the development team who are chartered to make the product better, and not as enemies attempting to cancel a program. Communication, a clear understanding of program requirements, and early resource planning are means that the test community can use to aid in maintaining a positive relationship with the PM.

Contractor Testing, Developmental Testing, and Operational Testing have some overlapping functions. DoD executes many overlapping tests throughout the product development cycle. The primary difference between similar tests is the controlling agency that is conducting the test. With limited resources and increased complexity of systems, a more integrated testing approach early in the product development cycle is necessary. An integrated approach will help facilitate earlier operational involvement. This integration must comply with statutory regulations. Under Title 10 U.S.C. 2399, “no person employed by the contractor of the system being tested may be involved in the conduct of the operational test and evaluations.” It further states, “A contractor that has participated in the development, production, or testing of a system for a Military Department or Defense Agency may not be involved in the establishment of criteria for data collection, performance assessment, or evaluation activities for the operational test and evaluation. “ (Stoddart, 2001, p.5) Recognizing the statutory limitations, a test team can develop a strategy that integrates the efforts of the contractor and development team and then the development and operational test teams. While restrictive, these regulations do not prohibit the interaction of the contractor and the operational community, rather they limit.

Independence of evaluation of test data is the essential element, not the taking of the data itself. Data should be available for all agencies to view and analyze. This can reduce test repetition, especially during the developmental portion of test. An understanding on data requirements for the various test agencies supports this recommendation. The environment where the data is collected must be meticulously recorded to ensure applicability for other test agencies. While data collected during DT cannot replace OT data, it can help support the OT effort if proper records are kept on its collection.

Response to perceived test "failures" is often inappropriate and counter productive. Similar to the commercial test philosophy, test failures should be viewed as learning opportunities and not program failures especially early in the development testing phase. A program office that understands this attribute will also have a better working relationship with its respective developmental test team. "Backing away from the pass/fail mentality and truly testing for learning," (Christie, 2004, speech) are philosophies supported by DOT&E.

### 2. Commercial T&E Best Practices

The DoD funded study conducted by Science Applications International Corporation (SAIC) examined high-powered companies with strengths in aviation, software, and technology and their approach to T&E. The developers of the study grouped their findings into four areas: test philosophy; test investment; test execution; and test evaluation. The list they produced in the document was extensive. Within the body of this text, applicable points considered relevant to this research are presented.

**Test Philosophy**

- Recognize that testing is a way to identify and solve problems early in the process in order to control time, cost and schedule late in the process.
- Increase T&E to assure product quality rather than reduce it to save T&E cost.

**Test Investment**

- Ensure early determination of the investment costs to acquire new capability for program support.

**Test Execution**

- Involve testers and evaluators very early: (1) ensure testers know test requirements; (2) ensure developers know requirements for test.

- Emphasis on concurrent and integrated T&E.
- Use measures and metrics.
- Train the in-house test workforce in test engineering disciplines.

**Test Evaluation**

- Correlate faults and solutions in a closed loop process to ensure problems are resolved.

(Science Applications International Corporation, 2002, pp. 3-4)

**T&E Best Practices Review**

**Test Philosophy**

- Recognize that testing is a way to identify and solve problems early in the process in order to control time, cost and schedule late in the process. A recurring theme throughout the research was that the PMs as well as Program Executive Officers (PEOs) should view the test process as a means to gain knowledge about the product they are charged to develop. The commercial industry, DOT&E, and the Defense Science Board have stated the importance of this T&E approach. Despite this support, PMs are reluctant to support a test strategy that increases a risk early in a program's development schedule. The benefit of identifying deficiencies early to afford time and the proper allocation of resources to effectively correct the shortfall does not outweigh the risk of spotlighting limitations of a program in the acquisition community. "The detection of a problem on an individual program makes that program vulnerable to criticism and possible loss of funding support." (GAO-98-123, 1998, p.16)
- Increase T&E to assure product quality rather than reduce it to save T&E cost. PMs should strongly resist the desire to reduce the scope of T&E to accommodate schedule slips or cost overruns. This scaled

back approach eventually leads to the identification of deficiencies during OT phases. Figure 7 conceptually illustrates a very real problem with reducing the test process. What may save in either cost or schedule today could cost in the future.

**"...and we can save 900 lira by not taking soil tests."**

Figure 7. Shortchanging T&E

**Test Investment**

- Ensure early determination of the investment costs to acquire new capability for program support. The resource requirements necessary to execute a test program require clear identification by the T&E planners and communication to the program office. Failure to do so will result in a catch up mode to gain funding for the resources. Including the test community early in the program's development and planning phases will help the PM understand the level of resource requirements. Involving the OT community early in this process is essential. Since funding for OT comes from the PM, it is essential to identify OT test needs early to ensure that there can be proper budget and resource planning. Involving the communities before a Milestone

B decision will ensure that the resource requirements are understood and properly recorded in Part V of the TEMP.

**Test Execution**

- Involve testers and evaluators very early: (1) ensure testers know test requirements; (2) ensure developers know requirements for test. Previous studies as well as leaders in the T&E field, such as the former DOT&E, Dr. Philip Coyle, have echoed this observation

  > Are you including the operational testers up front…They can help you early with requirements issues, with operational emphasis in the Request For Proposal (RFP), and with test and evaluation planning. Confronting such matters later will only increase costs and delay schedules, placing your program at unnecessary risk. (Coyle, 2000, p.5)

- Emphasize concurrent and integrated T&E. DoD is embracing this concept under the evolutionary acquisition approach. NAVAIR and COMOPTEVFOR are implementing such a strategy through a F/A-18 software upgrade program. Integration of testing will aid product development through the sharing of the limited resources from funding, range assets and support, and weapons asset availability. As noted in the earlier DSB discussion, integration T&E involving the use of the contractor will aid the government test communities. A positive blend of contractor and developmental T&E provides an opportunity to conduct early robust subsystem testing. This testing will enhance knowledge about the system components and provide an opportunity early in the development effort to correct deficiencies. This form of testing does not capture the eyes of others within the acquisition community and provides a platform to test for knowledge before the higher profile testing during DT and OT.

- Use measures and metrics. Establishing baselines or measures of test will allow the tester to effectively track the product development effort.

It will also afford the tester the opportunity to clearly communicate the testing progress to the PM.

- Train the in-house test workforce in test engineering disciplines. As resources are tight, it is very important as a PM to ensure that your T&E team is well trained and experienced. The knowledge they have will help increase the probability for correct planning and execution. In addition, "training provided to the program offices serves as a key agent in both creating a culture that is receptive to new practices and in providing the knowledge needed to implement new practices at the workplace." (GAO-99-206, 1999, p.2)

**Test Evaluation**

- Correlate faults and solutions in a closed loop process to ensure problems are resolved.
  - As a program progresses in testing, there is increased risk of overlooking or not resolving system failures or deficiencies. Establishing a clear approach to reporting, tracking, correcting, and verifying the correction will aid the product development process.
  - Within the ARM Program Office, there is an established approach, as illustrated in Figure 8, to handle the evaluation portion of test and track the observed faults or deficiencies for both hardware and software.

FAILURE REVIEW BOARD FLOWCHART

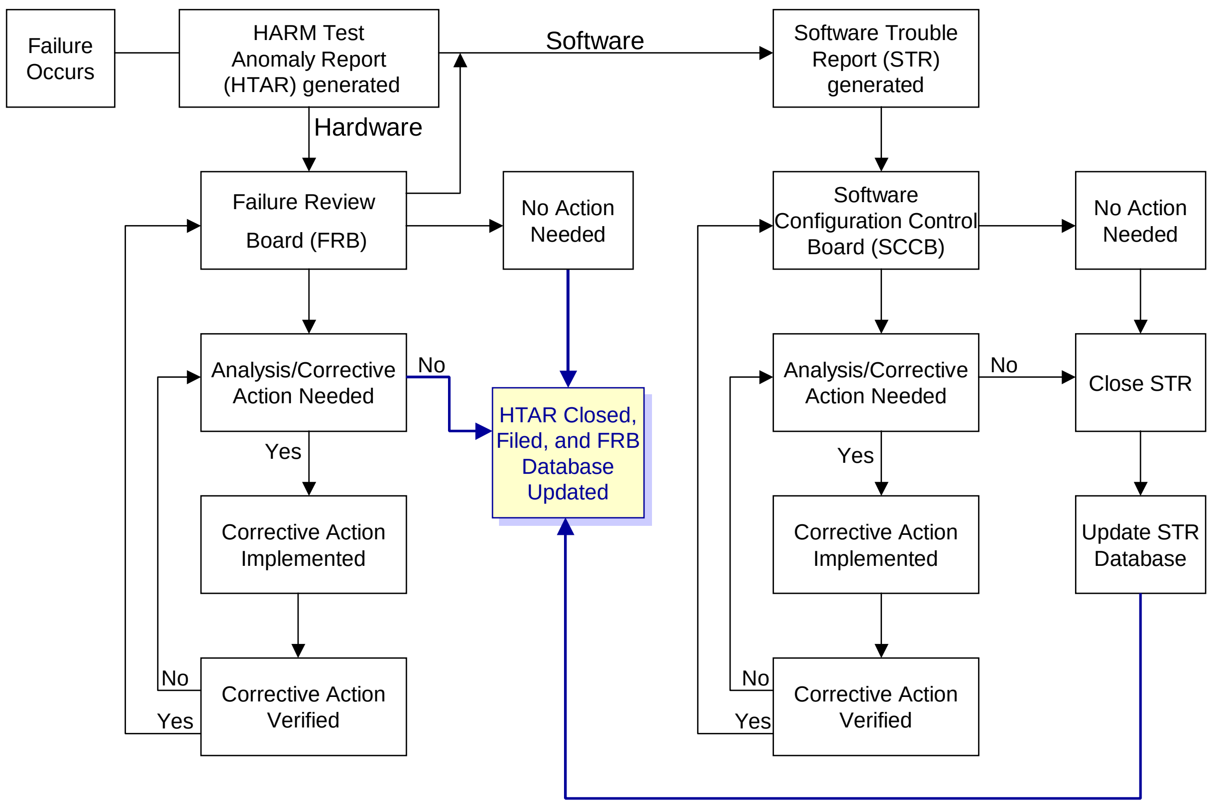


Figure 8. ARM Failure Analysis Chain

## G. SUMMARY

There are common trends identified by DOT&E that affect the ability of the PM to effectively test a system. These shortfalls are a product of DoD's infrastructure and philosophical approach to testing. The test infrastructure is slowly deteriorating. It has reached the point where the creation of a DoD Test and Resource Management Center was required. DTRMC, a recommendation by the DSB in 1999 and then by DOT&E in FY02, will be responsible to assess and allocate the necessary policies to rebuild the declining range infrastructure. Given a better set of tools, PMs will be able to more effectively and efficiently test a developing system at a potential cost savings. The reason is that with an improvement to the infrastructure, newer and more reliable range systems can be available.

While there is apparent progress in the range infrastructure shortfall, DoD philosophy towards testing still requires a transformational shift. While studies and recent DoD Directives have supported this shift towards the commercial best practice known as knowledge-based testing, the transformation has been slow to occur. There continues to remain a desire to minimize testing early in a program's development cycle. This process is exacerbated through the acceptance of high technology risks (low TRLs) at program inception. The commercial industry has learned through experience that a program cannot be successful if early knowledge of system capability is not attained. With the growing complexity of DoD weapon systems from an individual and interoperable perspective, DoD leadership must provide the necessary support to the PMs to support this knowledge-based approach. By doing so, it will foster an improved test culture as it will effectively allow the DT community to properly test the system.

THIS PAGE INTENTIONALLY LEFT BLANK

# III. LEARNING FROM HISTORY

## A. INTRODUCTION

Learning from the mistakes or good test approaches that other programs have used is good practice. In order to support this approach, respective programs should generate and make readily available documentation that discusses lessons learned. In the aviation community, specifically naval aviation, this is conducted through flight debriefs, squadron meetings, and professional magazines. During this time, aviators provide mission alibis. This offers others in attendance an opportunity to learn from the errors. By openly conducting this lesson learned feedback loop, there is a decreased risk that others will repeat the mistake and thereby increase mission success and/or save lives. In the acquisition community, there is very little opportunity to provide such feedback to the community. Fear of retribution or lack of time restricts wide dissemination of lessons learned from programs. Typically carried by word of mouth from experienced testers, this dissemination does not allow a large group to be educated. Furthermore, test teams are quickly disbanded at the end of a program before the production of a quality lesson learned document. There is also typically no support in the budget to provide such a document. The following section discusses some testing lessons learned from former programs.

## B. NAVAIR PERFORMANCE

A study was conducted by NAVAIR to evaluate trends and observations, shown in Figure 9, in the testing of systems sent to the OT community. The analysis covered 64 programs that were in OPEVAL or Follow-On Test & Evaluation (FOT&E) from FY-97 through FY-00. During this time, NAVAIR programs experienced a 3% failure rate in the area of operational effectiveness. While quite an impressive achievement, the associated operational suitability numbers were not as positive. In this area, 23% of the programs failed. (AIR 5.1E Brief, 2004) The results found that training, documentation, reliability, and logistic support were deficient.

NAVAIR

AIR-5.1E ACQUISITION T&E DEPARTMENT

**Common Suitability Issues**

**TRAINING**

- INCOMPLETE, NO TRAINING PLAN, NO TRAINER

**DOCUMENTATION**

- INCOMPLETE, INACCURATE, COMPLEX, AND/OR INADEQUATE

**RELIABILITY**

- RESULTS MAGNITUDES LOWER THAN THRESHOLDS (HARDWARE AND SOFTWARE)

**LOGISTIC SUPPORT**

- SUPPORT PLANS NOT AVAILABLE, COMPLETED OR FULLY IMPLEMENTED
- PARTS NOT AVAILABLE
- PARTS NOT IN NAVY SUPPLY SYSTEM
- MAINTENANCE LEVELS NOT IN PLACE
- NON-FLEET REPRESENTATIVE SUPPLY PROVIDED

Figure 9. NAVAIR Identified Common Suitability Issues (AIR 5.1E Brief, 2004)

This result supports the comment made by a former Commanding Officer for the Navy's Developmental Test Squadron, VX-31. When asked what he considered to be a deficiency in the way the Navy conducts developmental testing, he commented that the Navy does a great job testing the effectiveness of a system and identifying the goods and other aspects about a system's war fighting capability. Unfortunately, they generally fail to look at the entire spectrum of testing, which includes operational suitability. (Burris, 2004, interview) He further added that in his early days in the T&E community, he had been involved in a program that failed to confront this very issue. The program he referred to was the once highly classified Tri-Service Standoff Attack Missile (TSSAM) program. He commented that while the program was technologically mature, the reliability issues led to multiple firing failures. Each test failure was a result of different component failures. Schedule delays and increasing costs eventually resulted in program cancellation in December 1994. Cost growth had gone from $728,000 per unit in 1986 to $2,062,000 in 1994 (then year dollars). (Federation of American Scientists, 1998, ¶ 5).

In a study of Army Programs that included the ADATS (LOS-F-H) Air Defense System, Avenger (Pedestal Mounted Stinger), OH-58D (AHIP) Scout Helicopter and the Apache (AH-64) Attack Helicopter, suitability problems were also noted. In this study, recommendations to overcome the suitability shortfalls were provided. They include:

- Early Attention to Technical manuals resulted in a more accurate product and led to fewer logistics support problems during operational test;
- Technical manuals should always be a planned objective;
- Contractor technical writers should be brought to the training and testing locations to correct technical manuals as problems are noted by the users;
- All system training publications and manuals must be completed, reviewed, and selectively tested prior to the beginning of operational test;
- User experience and training before operational test is extremely valuable; and
- Training should be conducted at a proper point before operational assessment and should include prototypes and detailed mock-ups.

(Hoivik, 1997)

Similar to the Army study, the NAVAIR study provided some recommendations, shown in Figure 10, for future PMs and testers to consider as they execute a test program. The recommendations are the result of analyzing over 64 naval aviation programs of which 10 were ACAT I Programs.

NAVAIR

AIR-5.1E ACQUISITION T&E DEPARTMENT

**Recommendations**

- USE LOGISTIC SUPPORT REPRESENTATIVE OF FLEET CONDITIONS
- RELIABILITY WILL NOT GET BETTER IN OPEVAL -- ATTAIN LEVELS IN DT FIRST
- PROVE EFFECTIVE WORKAROUNDS BEFORE OPEVAL
- PROVE SOFTWARE MATURITY IN DT
- AVOID ENTERING OPEVAL/FOT&E WITHOUT PREVIOUS OT
- ENSURE RELIABILITY, DOCUMENTATION, TRAINING AND BUILT-IN TEST ARE READY
- HAVE OPEVAL LOGISTIC SUPPORT PLAN FULLY IMPLEMENTED
- HAVE TRAINING PLAN FULLY IMPLEMENTED
- IF ISSUES ARE IDENTIFIED PRIOR TO OPEVAL/FOT&E, SECURE A WAIVER
- ALLOW TIME FOR DOCUMENTATION TO BE DEVELOPED AND CHECKED BEFORE OPEVAL

Figure 10. NAVAIR Recommendations (AIR 5.1E Brief, 2004)

## C. LESSONS FROM OTHER PROGRAMS

Program test strategies, both past and current, offer a tremendous amount of learning opportunities for future programs. While each system presents unique challenges and is different in their mission and performance goals, their common successes and failures in the testing approach provide a basis for generic lessons learned.

### 1. Hubble Space Telescope

Launched in 1990, the Hubble Space Telescope, Figure 11, was a scientific effort that received worldwide attention. This attention resulted in global embarrassment as the first images produced by the telescope were out of focus. There had been an inherent flaw in the lens system. This flaw, detected on Earth six years earlier, was a result of ignored data by engineers and testers. National Aeronautic and Space Administration's (NASA) review determined that if the proper ground test procedures had been in place, along with the proper process to report and analyze the data, then this issue could have been rectified earlier, and at decreased cost and risk to the program. (Cotterman et al., 2000, p.127)

Figure 11. Hubble Space Telescope

Lessons learned:

- Data analysis;
- Marred failure analysis process; and
- Inadequate ground testing.

### 2. Standoff Land Attack Missile-Expanded Response Weapon

The Navy's upgrade to the Standoff Land Attack Missile (SLAM) is the SLAM-Expanded Response (SLAM-ER) program, Figure 12. Designed to address the Navy's requirements for a precision-guided Standoff Outside of Area Defense (SOAD) system, the weapon encountered continuous problems throughout its development and test cycle. During development, the program failed to account for historical deficiencies, specifically concerning data link interference and the resultant impact to the displayed image presented in the cockpit videos. As a result of this oversight, the maximum effective range from the target to facilitate weapon control was reduced.

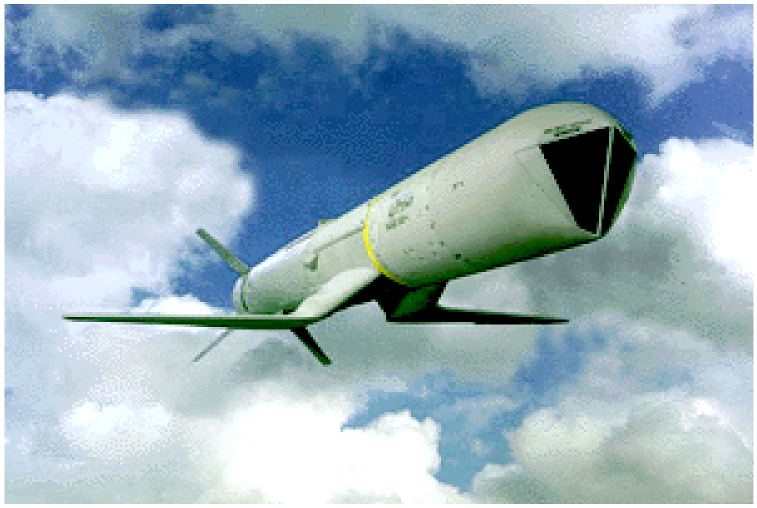

Figure 12. SLAM-ER Weapon System

Another factor that affected the program's ability to effectively test was the relationship between the PM and the testing community. The DT community, under much pressure from the PM, designed the tests to succeed rather than verify true system performance. They augmented the targets to make them easier to see through the weapon's seeker. They were testing for success and limiting knowledge gained. This created an adversarial PM to tester relationship. During the independent OT, without target augmentation, five of the eleven firings were a success. DT test expertise was also a contributing factor to the OT troubles.

> Test pilots and maintenance crews had become experts and intimately familiar with the test missiles. Thus, they knew how to work around problems, such as when the video images on the target acquisition system froze…test articles were prepared and maintained to be in the best condition. (GAO-00-199, 2000, p.40)

The compensation experienced in this program is a general concern throughout the test community. In a study conducted by LtCol Alford (USAF), where he evaluated the impact of test pilot compensation during aircraft acquisition programs, he stated,

> Test pilot compensation hides critical handling qualities cliffs that can lead to loss of an aircraft when encountered by less skilled

> pilots. This observation has vast ramification for test, evaluation, and development of all human interface systems. (Alford, 2004, p.23)

Further factors that affected the SLAM-ER program included the instability in system configuration. The system experienced configuration changes even in OT, thereby reducing confidence and knowledge gained. The SLAM-ER program also received unsatisfactory marks in operational suitability, specifically in reliability and maintainability. These scores were a direct result of the Mean Time Between Operational Maintenance Failure (MTBOMF) criterion not being met and poor built-in-test (BIT) performance. (AIR 5.1E Brief, 2004)

An additional contributing factor to the struggles that the SLAM-ER program experienced relates to their test approach. The program did not fully integrate an early operational assessment before proceeding to independent OT.

Lessons learned:

- Historical performance problems (feedback loop);
- Testing for success;
- Negative test culture;
- Lack of an early operational assessment;
- Performance compensation by test pilot and test maintainers;
- Reliability/maintainability issues; and
- Configuration stability.

### 3. F/A-22 Raptor

The F/A-22 Raptor, Figure 13, is an example of what can happen to a program's requirements as it begins to drag out for an extended period of time. Since development began in 1986, the advanced fighter aircraft's mission was to ensure future air-to-air dominance against the Soviet Union. As the threat changed and the development schedule for the aircraft extended so did the requirements. The USAF, in an attempt to save the program from the cutting floor, re-designated the aircraft from a single mission fighter to a dual role fighter/attack aircraft. This change in mission brought new mission requirements to a struggling development and test program. As of March 2004, GAO reported that these new requirements would require a budget increase of $11.7 billion. (GAO-04-597T, 2004, p.5)

Figure 13. F-22 Raptor

Additional factors have contributed to the slow development and test effort. The program managers and testers developed a very optimistic test strategy. They assumed that there would be no failures in hardware or software during ground testing. As a result they did not plan in their schedule any time to repeat or re-fly test events. They planned to always have an aircraft available for each scheduled test event, and they expected each event would provide productive information for the advancement of the program.

Test planners did not effectively evaluate the maturity level of the technologies incorporated in the aircraft. As a result, the development timeline extended reducing the available testing schedule and funding. Moreover, as the program experienced various test failures, the overall test program required reorganization leading to a curtailed approach. As an example, funding shortfalls required the elimination of range resource products, designed to support the test program.

Avionics testing was reduced to approximately half to save schedule and cost. To support this reduction, the program intended on combining multiple objectives to one test. This test approach has been negatively viewed by DOT&E.

> Never place your program at unnecessary risk by betting it on a single test…Any time you get into a situation where the outcome is going to be all or nothing, black or white, you probably need to rethink your test program. (Coyle, 2000, p.3)

Further reductions in testing will occur with respect to live fire air-to-air testing. Captive testing, rather than live fire testing, is now the intended strategy during IOT&E. Furthermore, production representative finishes to meet stealth specifications had not been flight tested before full-rate production. This increases the maintainability risk and is very similar to what the B-2 program experienced. (F-22 Raptor, 2001, ¶ 6)

Speed of testing is another identified weakness for the program. The Honorable Mr. Christie stated, “When the F-22 program fires but one missile a month in its test program, there is something profoundly wrong with the speed at which we can conduct testing.” (Christie, 2002, speech)

Recently there has been concern that the F/A-22 program will not meet operational suitability.

> The F/A-22 program is not meeting its requirements for a reliable aircraft, and it is not using a knowledge-based approach. The Air Force established reliability requirements to be achieved at the completion of development and at system maturity. As a measure of the system’s overall reliability, the Air Force established a requirement for 1.95-hours mean time between maintenance by the completion of development and 3-hours mean time between maintenance at system maturity. This measure of reliability represents the average flight time between maintenance actions. As of October 2003, the Air Force had only been able to demonstrate a reliability of about 0.5 flying hours between maintenance actions or about 26 percent of the development requirement and 17 percent of system maturity requirement. This has led to test aircraft spending more time than planned on the ground undergoing maintenance. (GAO-04-597T, 2004, p.8)

Lessons learned:

- Optimistic test planning;
- Immature technology;
- Data flow chain (slow test process);
- Addition of new requirements (air-to-ground);
- Stacking test events; and
- Suitability issues.

### 4. Theater High Altitude Area Defense Program

The THAAD, Figure 14, is a mobile ground based missile system designed to hit and destroy incoming ballistic missiles. This system along with the Patriot system complement each other with the THAAD working the higher altitude engagements and the Patriot engaging lower altitude systems. The program has experienced classic T&E problems.

Figure 14. THAAD Missile System

As the schedule slipped due to development problems, the program office began cutting test events. There was a reduction in ground testing events. This delayed identifying problems until flight test. "Several failures in flight tests of the THAAD system were traced to problems that could have been revealed in ground testing." (GAO-00-199, 2000, p.17) Shipping and the integration of many subcomponents occurred without the necessary ground test verification. The technology of the seeker was not mature enough to support the user needs, but due to schedule and the cost growth, the PM accepted the lesser technology and reduced the scope of the testing. (GAO-00-199, 2000, p.34) This approach did not afford the opportunity to establish early system knowledge. The immaturity of the seeker technology resulted in an unstable seeker configuration further hampering gaining knowledge about system performance. In addition to supporting a reduction in testing, they accepted a reduction in test instrumentation used on the missile system. This decision limited the test team's ability to evaluate missile system failures, which occurred during firings two through nine. The test community further developed a test plan strategy that was overly optimistic in hopes that technology would catch up throughout development.

During their review of the program, the GAO conducted multiple interviews with program officials concerning the troubled programs. Two very poignant comments were made concerning the test approach.

> Program officials acknowledged that they took many shortcuts in technology maturation, expecting to make up this knowledge during flight-testing. (GAO-00-199, 2000, p.34)
>
> According to program officials, the difficulty of the technology maturation process alone could not be accomplished in the time allotted. To satisfy the early fielding date, program managers opted to omit fundamental ground and subsystem tests and use flight-testing to discover whether the missile design would work. When the flight tests proved unsuccessful, the early fielding date was postponed and the requirement was eventually deleted entirely. (GAO-00-199, 2000, p.51)

Lessons learned:

- Insufficient ground testing;

- Limited instrumentation;
- Immature technology;
- Optimistic test planning; and
- Configuration instability.

**5. AGM-88D Precision Navigation Upgrade (HARM PNU)**

The HARM PNU, an international program, Figure 15, designed to incorporate an improved navigational software suit in the HARM weapon, increasing its geo-specificity, experienced cost and schedule overruns during the development effort. As a result, the DT&E effort was de-scoped. This de-scope led to a 44% reduction in flight test events. With a cut in flight tests, the T&E team began to increase the number of test events for each flight. The updated flight test schedule was success oriented and allowed for only two software updates with a projected four-week schedule impact. Flight test events were scheduled approximately 10 days apart. While executable, this schedule did not account for the actual 10-14 day data analysis process that would occur after each flight test event. The original plan allowed for only a 3-4 day data turnaround time. The international complexities and total data package size prevented smooth data transfers between the companies involved in the development effort. On one occasion, to reduce the data flow chain, the test team sent a US engineer to a partner country to deliver the data from a flight test.

The DT strategy did not capitalize on the important lessons learned from the Contractor T&E (CT&E) phase. CT&E took six months longer than anticipated as a result of data analysis problems, flight test planning problems, data exchange delays, incomplete aircraft and weapon’s integration software, and subsystem interface problems. Program management incorrectly concluded that since the CT&E phase took longer, there was greater knowledge gained about the maturity of the system, and as a result, the DT effort would go smoother. During the DT test period, PM pressure necessitated testing to support schedules rather than when the system was ready. This approach resulted in many flight test events that did not increase the knowledge base to support the development effort. In addition, it also created a negative test environment.

Figure 15. HARM Missile

Other test strategy failures occurred with the integration of the navigational software and navigational hardware. Each product, developed in a separate country, did not undergo subsystem integration testing before shipping to the US for full system testing. The result of this stovepipe effort proved to be the downfall for the missile program. The hardware and software systems did not function properly when integrated. The failure to perform integrated subsystem ground testing, during the development of this software and hardware, resulted in insidious navigational problems. As a result, the system could not achieve the system specifications and operational requirements, and eventually resulted in the conclusion of the program. The test strategy also did not support an early operational assessment. There was a plan to perform operational scenario testing for the last two firings, but with the program delays and slide in test schedule, these firings were re-scoped to support the constantly changing software builds. By the conclusion of the program, many suitability issues had not received the attention necessary due to the effectiveness problems. Key suitability areas that were still deficient were documentation, training, and reliability.

Lessons Learned:

- Optimistic test planning (stacking test events);
- Data flow chain;
- Insufficient subsystems testing;
- Inefficient ground testing;
- Suitability issues; and
- Schedule-driven rather than event-driven.

### 6. DarkStar Unmanned Air Vehicle

The DarkStar Program, Figure 16, was an ACTD designed to demonstrate the military utility of the unmanned aircraft. Originally scheduled as a two-year program, it suffered a major setback after the aircraft crashed during its second mission. The causes for the failures were a direct result of a poor program and testing strategy. The termination of the program came as result of reduced funding support. Cost and schedule growth had increased more than 100%.

The program was marred for a variety of reasons.

> The DarkStar's components and subsystems were not adequately validated before flight testing began. PMs curtailed some testing earlier in the program to stay on schedule. Limited knowledge about the aircraft's performance contributed to the crash of the first test vehicle. For example, the fuel system was not sufficiently instrumented or ground tested before flight tests began. Some key sensor testing was deferred until after flight-testing. Also, the contractor made extensive use of commercial components without testing or qualifying them for use on military systems…To save money, managers decided not to construct an "iron bird", which is a physical replica of the aircraft's hydraulics and mechanical

> subsystems…Problems surfaced during the first flight test that were not fully investigated and resolved due to time constraints. Braking and flight dynamics problems were not resolved prior to the next flight which resulted in a catastrophic failure. (GAO-00-199, 2000, p.37)

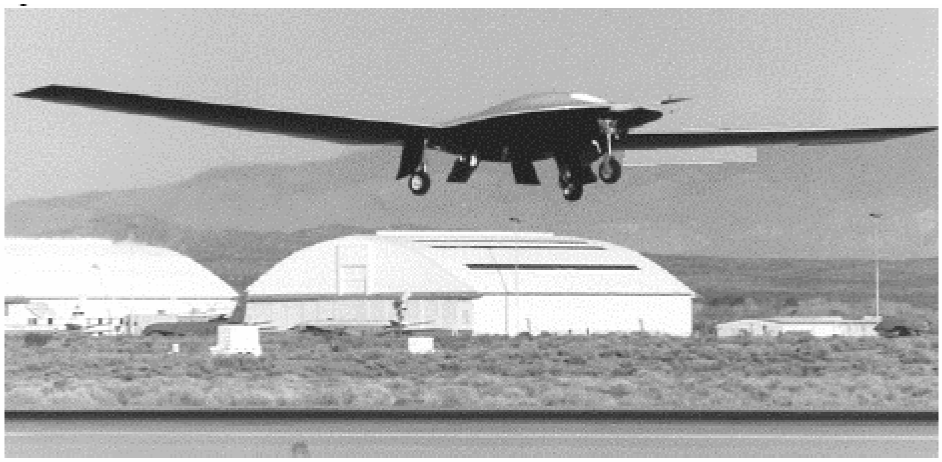

Figure 16. DarkStar Unmanned Aerial Vehicle

Lessons learned:

- Limited instrumentation;
- Insufficient ground testing;
- Inadequate prototyping; and
- Marred failure analysis process;

### 7. B-2 Stealth Bomber

While the B-2 Stealth Bomber, Figure 17, is combat proven, the testing community failed to fully test the system, and as a result, the USAF must deal with some cost drivers with respect to aircraft suitability. Effectiveness testing was successful, but non-operationally representative environments provided the basis of most flight tests. These tests, conducted in good weather, masked the true maintainability problems. Exposure of these problems did not occur until after Initial Operational Capability (IOC) and post full-rate production. Post IOC, B-2s returned from training missions with damaged skins, reducing their stealthy characteristics. "End result is that for every one hour of flying it takes 45 maintenance hours to fix, on average. Essentially only 33 percent of the aircraft can fly at one time." (Umansky, 2001, ¶ 11) This is a lesson that the USAF should consider with their F/A-22 aircraft.

Figure 17. B-2 Stealth Bomber

Lessons learned:

- Non-operational test scenarios; and
- Suitability testing.

### 8. AIM-9X Program

The AIM-9X Program, Figure 18, was a joint A/A weapon program that was designed to provide the aircrew with an off bore sight capability in the short-range air-to-air arena. The AIM-9X Program was a major upgrade to the existing weapon system. The program had its struggles, but the former integrated product team lead said in an interview that the test strategy used poised it to handle the challenges that it faced. Production Representative Missiles (PRM) went to the OT community early to build hours and support operational suitability. The OT community flew the weapons even if the missile was not part of a test event. Exposure of the system to many unscripted operational test events occurred during this time, affording the opportunity for excellent operator feedback early in the program's development effort. This feedback was then folded into the weapon system, which greatly enhanced the final product.

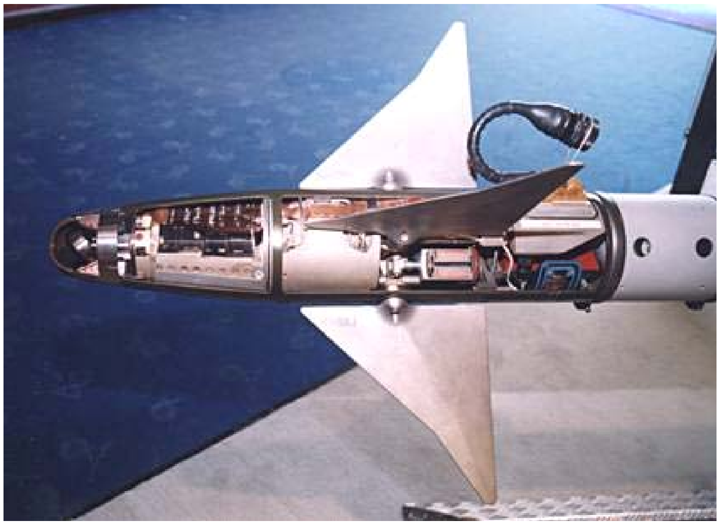

Figure 18. AIM-9X Seeker

The early involvement by the OT community further aided the development effort to meet operational suitability. The OT maintainers recognized a major flaw in the weapon's storage container. This resulted in an early modification to the container. Failure to capture this information before independent OT would have resulted in program delays. Another issue that was handled early in the program's development was the safe-and-arm handle. Operational maintainers recognized a flaw in the design. This flaw, had it not been corrected, would have prevented them from operating the handle with cold weather or chemical weapons gear. (Converse, 2004, interview)

The data flow chain, specifically the data analysis portion, took longer than desired. Because of the multiple agencies involved in the testing, there was a long turnaround time during the envelope expansion phase. There were four different agencies in four different locations responsible for evaluating the data.

Despite some foresight in test planning, the program did suffer because of range support issues. OPEVAL, despite early OT involvement, dragged out due to airborne target problems. This resulted in the AIM-9X program delaying live-fire tests until QF-4 aerial targets became available. This delay in aerial target availability highlighted a range infrastructure concern sited by DOT&E. Targets, whether airborne, sea based, or land targets, are becoming more difficult to acquire. Systems are becoming more precise and advanced in their ability to identify and track a target. In some cases, they can discriminate based on actual target appearance. Ranges do not have the funds to fully support higher fidelity target requirements to account for the advances in weapon capability. (DOT&E, 2004, p. 343) As a result, the financial impact that is incurred for higher target fidelity is typically transferred to the programs, which are already under funded. This drives a PM to push for less field testing.

Lessons learned:

- PRM early in the hands of the OT community can be beneficial;
- OT maintainer involvement early will help in suitability compliance;
- Data flow chain needs to be efficient; and
- Availability of representative targets.

#### 9. Tactical Tomahawk

The recent Tactical Tomahawk Weapon System (TTWS) test program, Figure 19, received unsatisfactory OT scores in the area of suitability. The areas of concern were documentation and training. The highlights from the overview stated that the training was insufficient to support operations related to the upgrade of the missile system. The documentation was unsatisfactory due to missing information or incorrect documentation. (Duarte – TTWS PM, personal communication, June 4, 2004) Ironically, the program also received unsatisfactory scores in suitability during another upgrade effort in FY97. During that test program, documentation, human factors, and reliability issues resulted in unsatisfactory test scores. (AIR 1.6 Brief, 2000)

The Tomahawk missile has proven its value in strikes in Bosnia, in Desert Storm, and in Desert Strike. A new version proposed by the Navy, will greatly increase the missile's capability yet result in half the cost of a new production run. U.S. Navy photo.

Figure 19. Tomahawk

Lessons learned:

- Unsatisfactory documentation;
- Inadequate training; and
- Not learning from historical performance.

### 10. High Mobility Trailer

Systems of all sizes can experience some of the most common problems. The Army High Mobility Trailer, Figure 20, developed in 1993, failed operational use. The truck trailers encountered serious safety problems and damaged the trucks that were towing them. The Army was required to purchase and then modify at a substantial cost. Analysis of this program failure indicated that the Army never conducted a field (operational) test before procurement. (GAO-00-15, 1999, p.6)

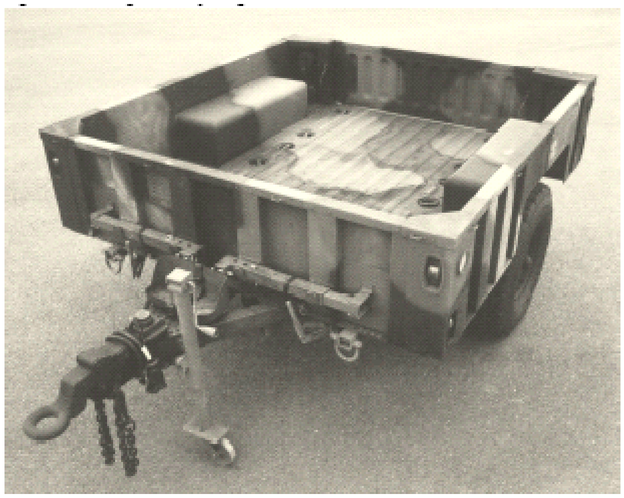

Figure 20. High Mobility Trailer

Lesson learned:

- Non-operational testing.

## D. ACADEMIC INSTRUCTION

Timeless problems plague testers as they evaluate a weapon system. Figure 21 lists some common areas of lessons learned from multiple Army programs. Created in 1996, the inclusion of this dated information is to highlight the commonality of problems between the services and indifference of time with respect to the issues DoD faces in T&E.

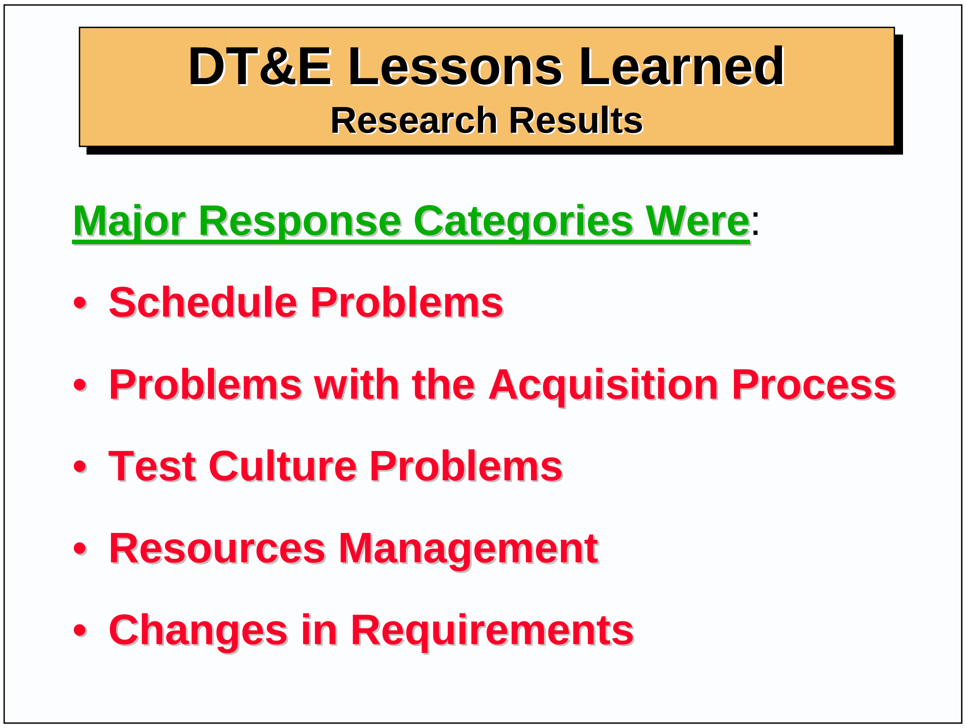


Figure 21. DT&E Lessons Learned

(Hoivik,1997)

## E. REQUIREMENTS

### 1. System Requirements

Clearly communicating system requirements, Figure 22, and Concept of Operations (CONOPS) to the entire test team will result in a system being properly tested.

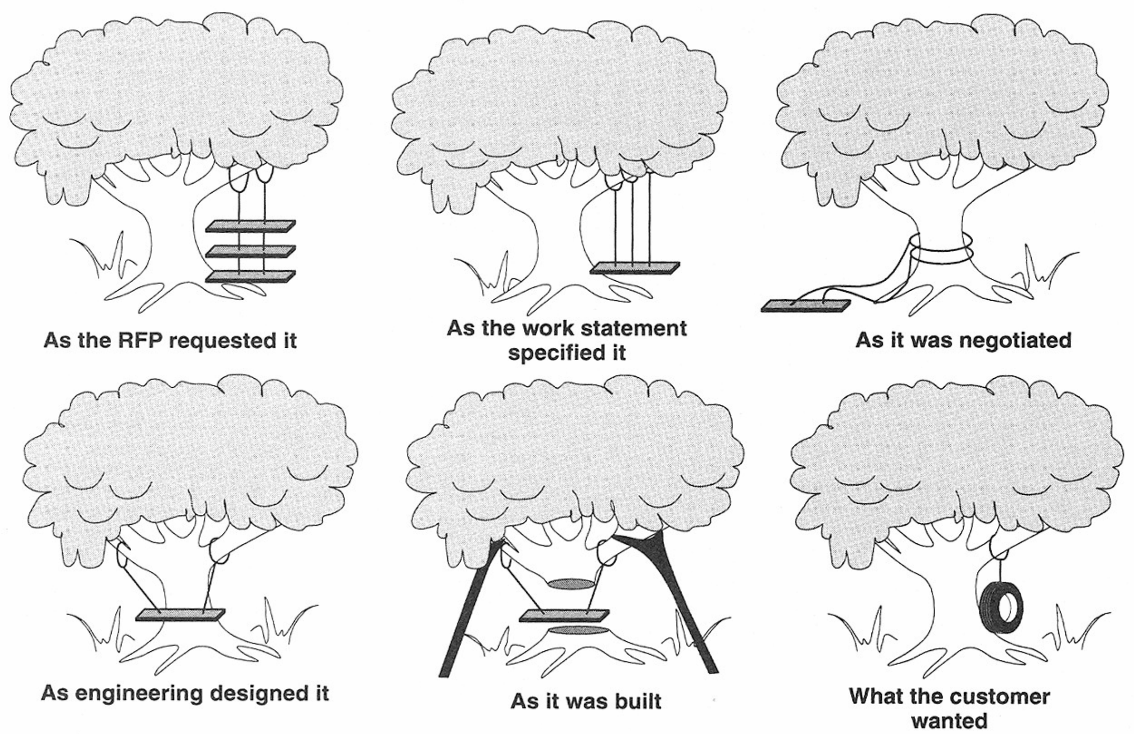


Figure 22. Views of Swing CONOPS
(Cotterman et al., 2000, p.113)

Although simple, this illustration highlights the importance of clearly defining and communicating the requirements. System requirements can be user requirements, functional/capability requirements, and performance requirements. They are then flowed into design or system specifications. Demonstrating specification compliance, while one of the DT community's test functions, does not guarantee compliance with the operational requirement, as shown in Figure 23.

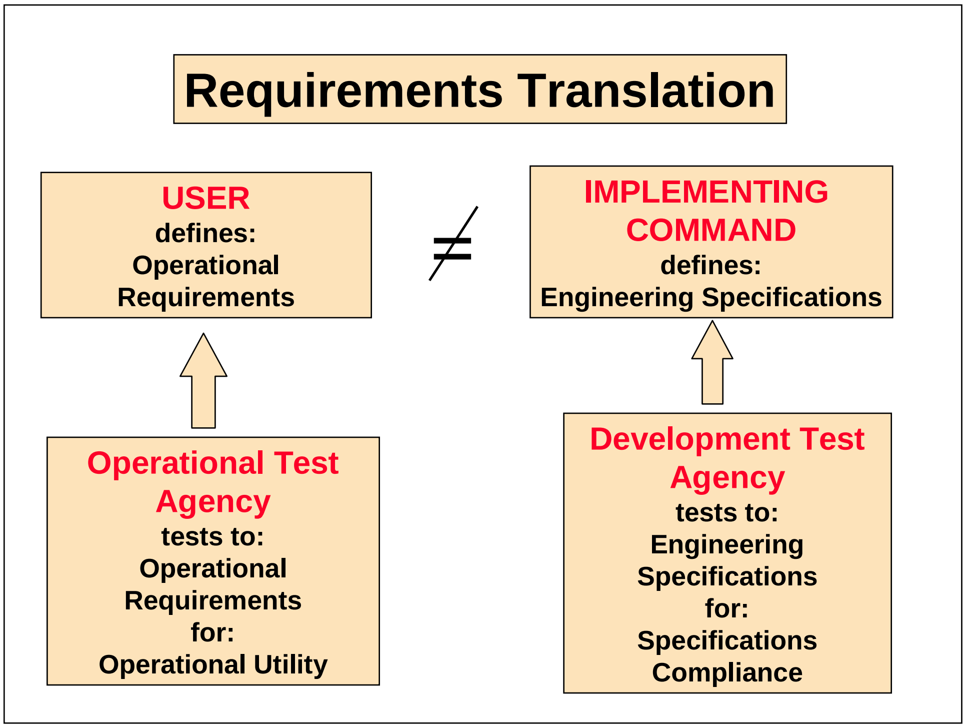


Figure 23. Requirements Relationship

(Owen, 2004)

Early involvement by the OT community and continuous communication between the DT and OT communities will reduce the risk of the DT community performing tests that do not effectively evaluate the system. Multiple sources of documentation capture the requirements and operational needs to support system development, Figure 24. Under the new acquisition guidelines, the Initial Capabilities Document (ICD) is equivalent to the Mission Need Statement (MNS), and the Capabilities Description Document (CDD) and Capabilities Production Document (CPD) replaced the ORD.

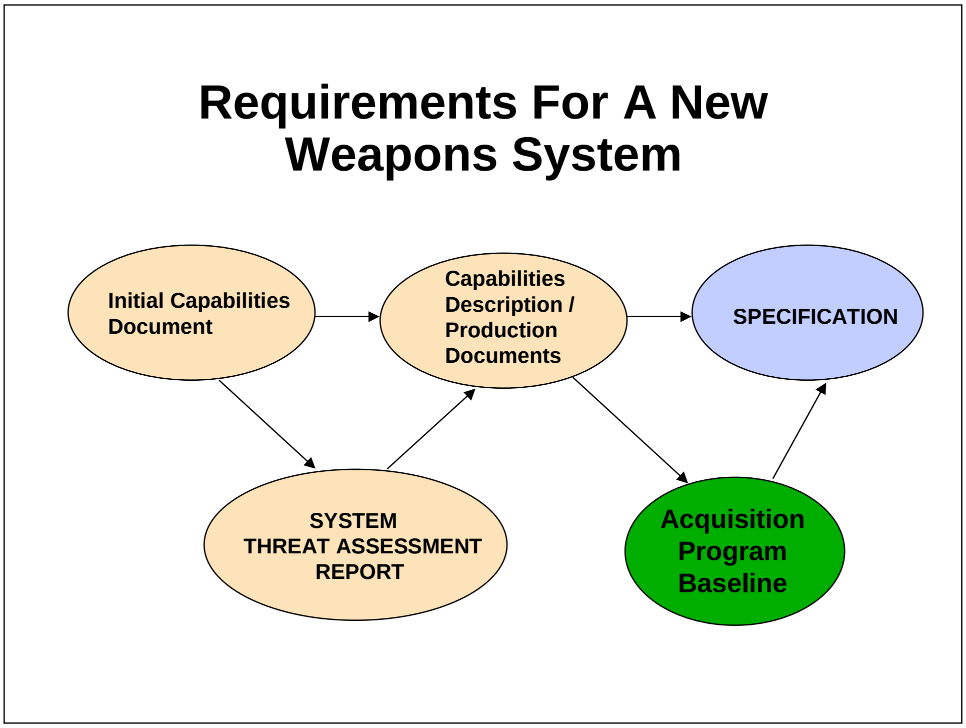


Figure 24. Requirements Documentation Sources

Figure 25 illustrates where the key acquisition documents fit into the acquisition process. The figure further shows the evolutionary acquisition process and the incorporation of the increment builds during the product development process. In an incremental process, "a desired capability is identified, an end-state requirement is known, and that requirement is met over time by development of several increments, each dependent on available mature technology." (Wascavage, 2004) This acquisition process allows the testing community to test and evaluate to a specified level of capability for each respective increment. While the overall system requirements do not evolve, the PM, user, and tester understanding of the incremental capabilities that will be introduced at various periods is essential to effectively establish the test strategy. Demonstration of controlling and testing to incremental requirements occurred during a major aircraft development effort.

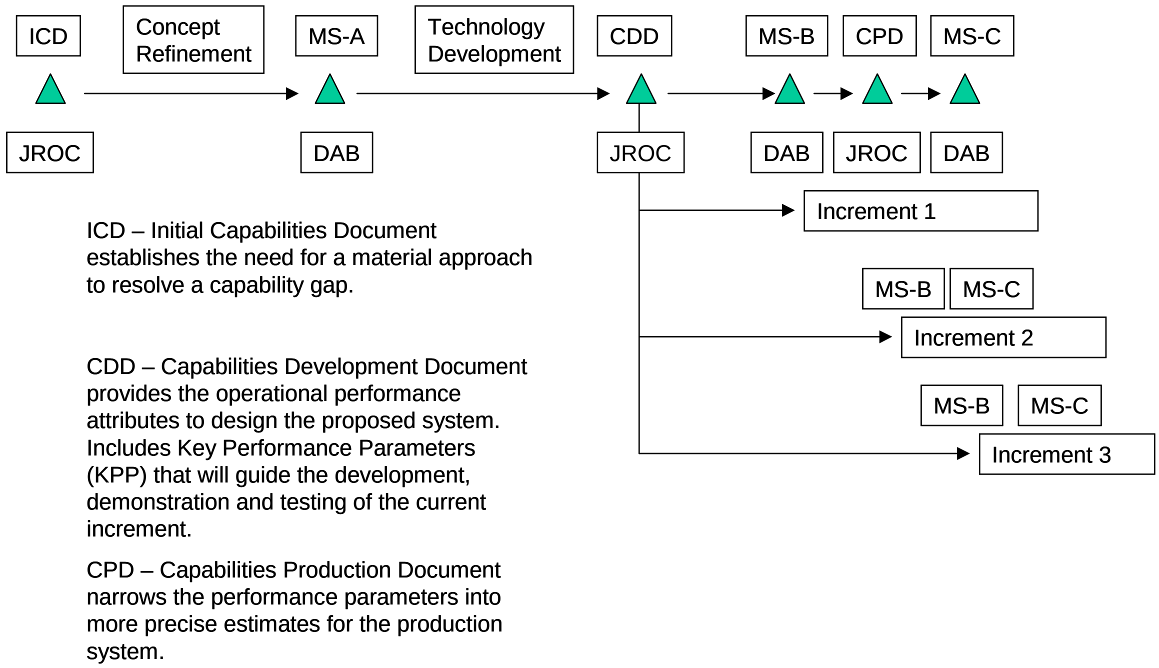


Figure 25. Requirements and Acquisition Process

(DoDI 5000.2, 2003, p.3)

During the testing for the F/A-18E/F Super Hornet, there was a strong desire to show and test the full capability of this weapon system. Testers overcame the PM and personal pressures to test beyond the scope of the initial test effort. Built upon a strong evolutionary development process, the program incorporated FOT&E into its test strategy to ensure continued testing after demonstrating basic combat capability. In an interview with CAPT Steve Burris, F/A-18 Advanced Weapons Lab (AWL) military lead during the testing, he stated that there was a “box” that the platform was originally required to perform to for the first test phase. The team continuously reminded each other not to stray into putting more in the box than what was required. One example of staying inside the “box” involved the testing of only 10-weapon load-out configurations. While hundreds of possible weapon load-outs exist, the Fleet user, through an operational executive committee, picked their top 10 weapon load-outs. These weapon load-outs defined the first phase of testing. Now after Initial Operational Capability (IOC), the aircraft continues to qualify other configurations beyond the top 10. Keeping the requirements focus for a particular phase of test inside the "box" proved to be a successful test approach as the aircraft met the deadline for first deployment.

While the F/A-18E/F development effort was successful, product development timelines for DoD generally run behind schedule, and as a result, CONOPS or the threat assessment can change before fielding a system. Therefore, there is a tendency to add to the basic requirements of a system to maintain its viability. This can create a cascading effect, Figure 26, potentially leading to new system problems, increased cost and schedule, and reduced system performance.

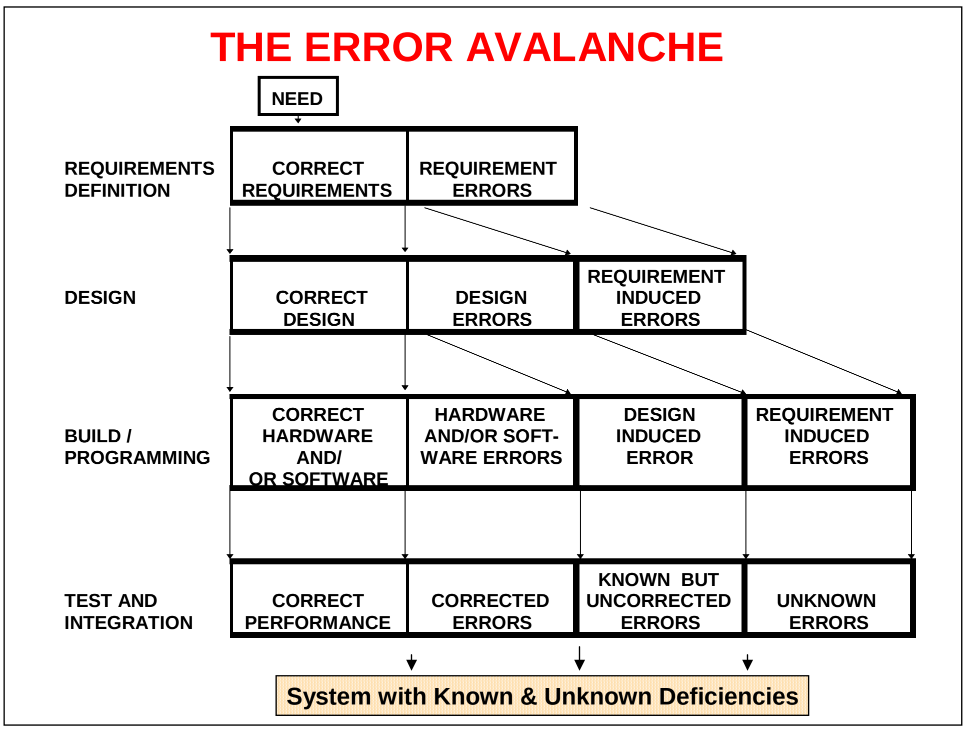


Figure 26. Requirements Instability

(Owen, 2004)

An example of not maintaining requirements is the Navy's Radio Frequency Countermeasure System. This system, designed to work on an F/A-18E/F, met the initial requirements for the Navy. The scope was a five-year development effort. Before start, during a joint review, the USAF determined they wanted to participate in this program and as a result instituted a new set of requirements that were more demanding than the Navy requirements. DoD did not approach this change from an evolutionary position and attempted to combine both Services' requirements for the first build. The late USAF requirements precipitated a program slip. (GAO-01-288, 2001, p.32) As a result, the F/A-18E/F was not able to deploy with the intended EW system and eventually deployed with a less effective system. The original system is still in development seven years after inception.

While the concept of requirements "creep" is not one that should be encouraged, there are times that altering or accepting a change is best for the program. "That combination—increased understanding and improved technology—often leads to the conclusion that the system's requirements need to be changed or expanded. This is not a bad thing…" (Ward, 2003, p.32) Reaffirmed by VADM Bennitt (Ret) in an article presented in the ITEA Journal, he stated,

> But any effort to adjust requirement parameters is generally viewed as a program failure, or as an effort to circumvent the acquisition system, rather than as an acknowledgement of the capability that is realistically achievable within an expected timeframe. Hard objective data must be provided to support a program manager's decision to deliver timely upgrades to the Warfighter, even if that means moving higher-risk capabilities into a later development spiral. (Bennitt, 2004, p. 7)

The PM and the DT test community must evaluate any requirements changes that are made throughout the development cycle and only under the strictest guidelines accept a change/alteration to the requirements. Requirement changes require an impact assessment to the test effort. If there is a significant improvement and the test resources are available to support the change with minimal program impact, then an alteration may be acceptable.

### 2. Special Test Requirements

Test requirements also present some unique challenges. Throughout a test program, various agencies require different information from a specific test event. A test strategy should consider what the data requirements are for each interested test agency. As a tester, planning for the multitude of test data requirements early in a developing program will prove very beneficial throughout the entire test evolution, even during the OPEVAL period.

> Engineers and testers should work together early on to ensure that key components are easily instrumented or readily provide necessary test data. In some cases, this is simply a matter of approaching the development with testing in mind. In other cases, creative methods may be required. Progress should be aided by the fact that, as information technology becomes more available and pervasive in systems, the ability to collect, export, and analyze test data will dramatically improve. (Sega, 2003, p.7)

Instrumentation data requirements are generally the concern of the DT community, as the OT community typically does not allow instrumentation gear installed within their systems. The operational community is interested in testing the production representative system, as delivered to the Fleet/Field. With the combination of reduced resources, increased test complexity, and growing desire to integrate DT and OT testing, vigilance by the DT community to understand the instrumentation needs for the OT community early in the test-planning phase is necessary to avoid delays to test. The Army TACMS testing in 1990 experienced this problem. Before the start of IOT&E, the OT community desired to have their Fleet representative launchers instrumented. The contractor for the launcher refused to perform the request, and the OT community was required to hire and independent contractor to design and install instrumentation. This eventually resulted in a test delay of two weeks. A lesson learned from this event was pointed out in the after action report. “Additional instrumentation of the systems, if needed at all, must be completed before start of test to avoid delays; might be best to plan for test instrumentation in the design of the system.” (Dillard, 1990, p.57) In a draft white paper, the author of the after action report and participant in the tests reflects on this event and the importance it has on the testing strategy. “Instrumentation is the single most important consideration that our Block I program has neglected in development.” (Dillard, 2004, ¶ 16)

## F. UNDERSTAND THE OT PLAYBOOK

During the building of the DT strategy, the DT community should evaluate the type of testing expected by the OT community. Understanding the OT strategy will help the DT community create a more robust test effort. The DT team can gain this knowledge via the Service-unique Operation Test Director (OTD) Guidebook. In the Navy, this document clearly identifies the operational community's strategy in planning and executing their tests. The document discusses two primary areas for program evaluation—effectiveness and suitability.

Operational effectiveness is the overall degree of mission accomplishment of a system when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, tactics, survivability, vulnerability, and threat. (COMOPTEVFORINST 3960.1, 2004, p. G-8) The building blocks to operational effectiveness are shown in Figure 27. The data and observations collected during testing are compared against the Measures of Effectiveness (MOE) and the Critical Operational Issues (COI), which are defined in the TEMP. COIs are defined as, “the critical aspects of a system's operational effectiveness and operational suitability that are intended for resolution during OT. (COMOPTEVFORINST 3960.1, 2004, p. G-3) MOEs are defined as, “a measure of operational success that must be closely related to the objective of the mission or operation being evaluated, for example, kills per shot, probability of kill, effective range, etc. A meaningful MOE must be quantifiable and a measure to what degree the real objective is achieved.” (Helm, 2002, p.10) These test metrics are developed during the test planning process and are the basis of the evaluation during OT.

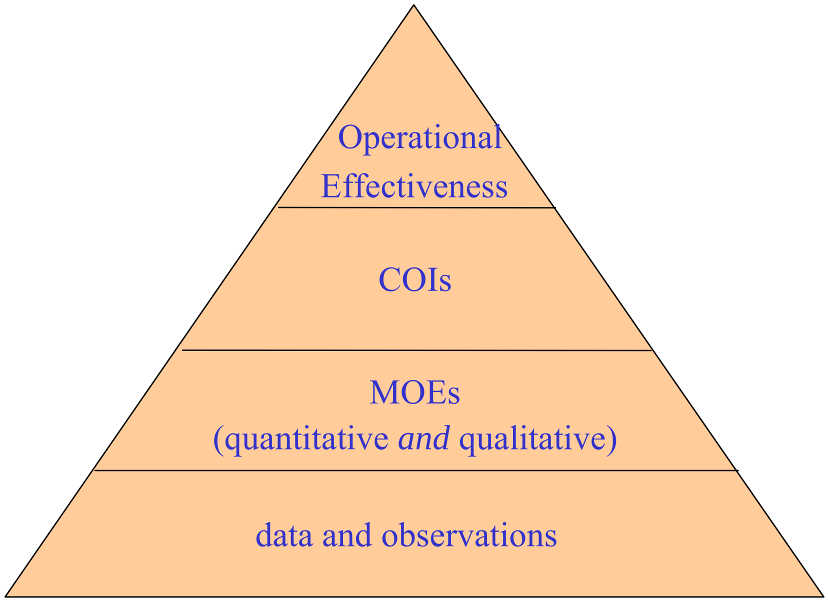


Figure 27. Operational Effectiveness Building Blocks

Operational suitability, highlighted in earlier sections, is another area evaluated by the OT community in the assessment of a system that has been an area of neglect in a product's development effort. The OTD Guidebook defines a list of the suitability requirements.

- Reliability
- Maintainability
- Availability
- Logistic supportability
- Compatibility
- Interoperability
- Training
- Human Factors
- Safety
- Documentation
- Transportability
- Wartime Usage Rates
- Manning Requirements

- Natural and Environmental effects and impacts

Operational suitability is the degree to which a system can be placed satisfactorily in field use with consideration given to the criteria listed above. (COMOPTEVFORINST 3960.1, 2004, p. G-9)

The OTD Guidebook also provides information concerning the two types of testing that the OT community can choose. They are scenario-oriented or operation-oriented testing. (COMOPTEVFORINST 3960.1, 2004, pp.6-7) Because the operational use of weapon systems changes with the changing operational situation in the battlefield, the OT community tends to test in a scenario oriented testing environment. In this type of testing, the OT community will develop scenarios that meet or match current tactics and procedures, as well as present a realistic threat environment. Early determination of unique OT resource requirements, because of scenario testing, can be captured in the TEMP if the DT community plans early with the OT community. Early planning can increase the probability that the correct resources will be available to support the OT effort. Furthermore, the DT community can understand the OT strategy and test to a similar level at the appropriate time in the development phase. This will increase system performance knowledge when in an OT representative test environment. As pointed out earlier, gaining early knowledge about system performance, similar to the commercial industry practice, can help reduce program risk.

## G. DATA SUPPORT / ANALYSIS

Throughout a test program the desire for data is unquenchable. Engineers, managers, and testers look for the information that the data streams and charts produce. This valuable information offers insight into the health of a program. With the reduction in the range infrastructure due to budgetary cuts, this necessary requirement has proven costly when not available. “Obsolete facilities and equipment increasingly fall short of data collection requirements.” (Gehrig et al., 2002, p.58) telemetry (TM) needs, data turnaround time, availability of range instrumentation systems, and the available frequency bandwidth all play important roles in data collection and analysis.

Reducing the time in the data flow chain will aid in increasing the tempo of testing, which is one of the recommendations from DOT&E. By reducing the development testing timeline through shortened data flow processes, a reduction in the overall product development timeline is achievable supporting Secretary of Defense Rumsfeld's desire to "minimize development time." (Dulin, 2001, p.75) Figure 28 presents this author's view of the data flow chain.

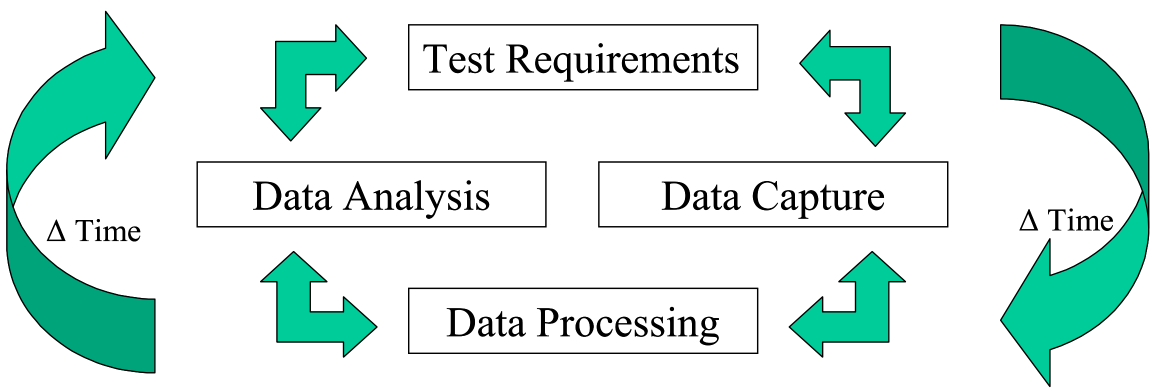


Figure 28. Data Flow Chain

Identifying the respective testable requirements is the first part of the flow chain. Once identified, the test team determines the best means to test and capture the data. This data capture can be from telemetry streaming information, video displays, developmental tester notes, or some other medium. Once captured, a test team must determine the process to proceed to the next test event. This process can be a real-time voice report from the tester, or it may involve time-consuming engineering reviews. The final phase of the flow chain involves analysis. This is the phase where the test team confirms the results, especially if poor performance or unexpected results were observed and a root cause must be discovered. For a test team, the strategy should be to reduce the overall time it takes to complete the data flow chain. Highlighted in this research were two programs that demonstrated both extremes to the data flow chain process. For one of the contractors in the JSF program, a mobile and self-contained TM collection van was used to allow real-time processing and analysis. This afforded stepping through test events in a rapid fashion, thereby maximizing the available test time. On the other end of the data flow spectrum was the HARM PNU program. This international program did not have an efficient data process after data capture and was plagued with a long data analysis process.

Understanding this data flow chain and working to minimize the timeline will become more important as weapon system complexity and the amount of data continues to grow. Poor data analysis processes will slow the T&E schedule and could result in very little being learned from each test. As VADM Bennitt (Ret) states,

> …the tyranny of the data avalanche. The F/A-22 and the F/A-18E/F development programs pushed the envelope with regard to the challenges of gathering and analyzing massive amounts of data. Data will be acquired by several test articles, operated at multiple sites. Bring on Joint Strike Fighter (JSF)! Fourteen test aircraft, four distinct "customers," multiple sites, highly instrumented aircraft and engines and growing international participation-the data "take" will be counted in hundreds of terabytes. The test community must be focused, disciplined and fully integrated into every aspect of the development process or it runs the risk of drowning in the data. (Bennitt, 2004, p. 7)

## H. INTEGRATED T&E

Throughout the research, there have been numerous references about involving the operational community earlier in the test process. Former DOT&E Dr. Coyle presented the question, "Are you including the Operational Testers up front?" during a speech in 2000 at Fort Belvoir. (Coyle, 2000, p.5) This can prove to be very difficult but the perseverance by the DT team to involve and inform the OT community on test decisions can save program time and money. NAVAIR recognizes the importance of integrating T&E. As a result, the initiation of a pilot program with the latest F/A-18 software upgrade effort has begun. The 19C Operational Flight Program will develop an integrated test plan with a test strategy to use Integrated Test and Evaluation (IT&E). The primary objective is to efficiently execute software upgrade testing with a smart test planning approach between the DT and OT communities. Figure 29 is a general representation of the level of effort that will be expected from the respective agencies throughout testing. Highlights to the IT&E concept include sharing of test and range assets, developing a common test plan, and conducting a shorter independent operational evaluation. Within the T&E philosophy, there will be a minimization of repeat test events, thereby accelerating product development, identifying and correcting potential operational problems earlier in the development and test cycle and ultimately saving program cost.

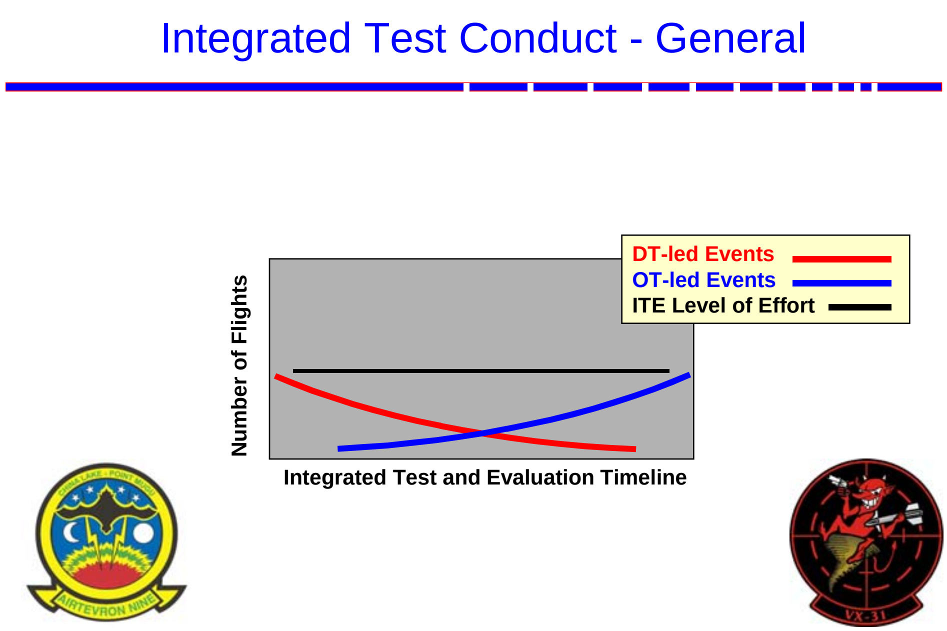


Figure 29. Integrated Test Conduct for Each Agency

Figure 30 shows a lower functional level with increased detail regarding the responsibilities and types of tests for major test phases. Parallel DT and OT efforts will occur to support this integrated approach. The complexity of the tests begin to increase as the program matures, and it is during this time that the OT level of effort increases. The overall objective is to support:

- Quicker test process;
- Capability-based testing; and
- Consolidation of limited resources.

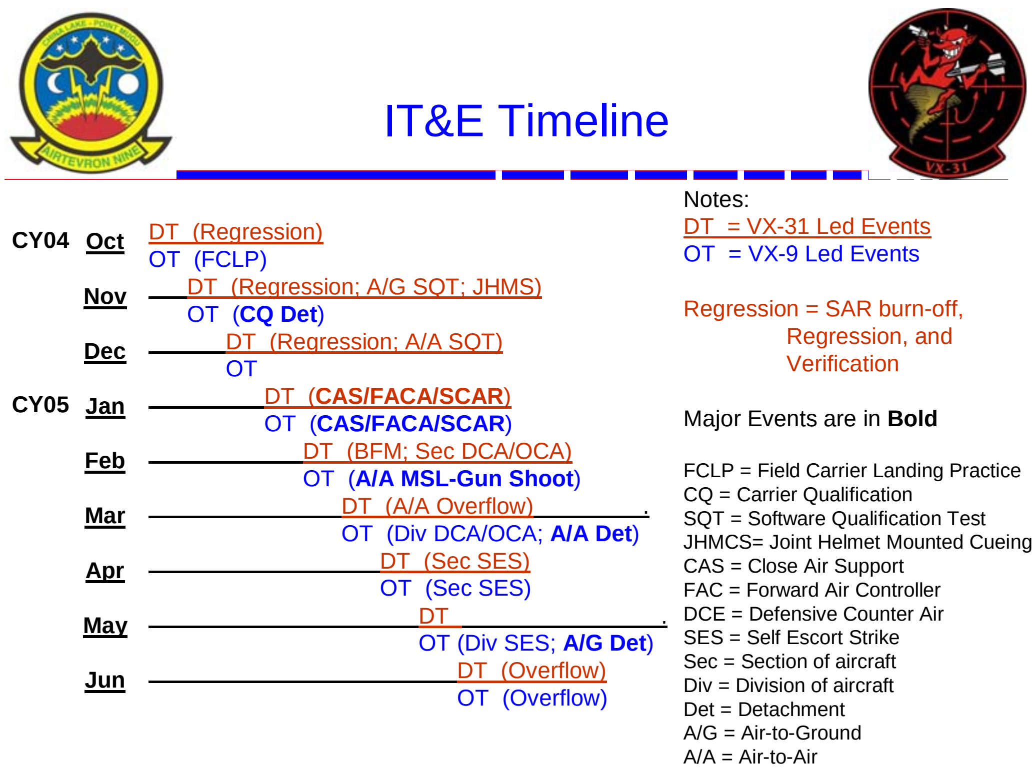


Figure 30. IT&E Level of Effort Timeline

## I. SUMMARY

The challenges that face DoD in testing are not new. The quick review of 10 programs in this research spotlights recurring themes in DoD testing:

- Insufficient ground testing;
- Insufficient system instrumentation and data analysis;
- Testing in non-operationally representative environments;
- Hardware and software configuration instability;

- Very little focus in operational suitability issues;
- Optimistic test planning;
- Schedule driven testing rather than event driven testing; and
- Negative test culture.

As a DT team begins the process of developing a strategy to effectively test a product, they must understand these historical problems. This will afford them an opportunity to mitigate the probability of repeating the same mistakes.

Testers should also understand some of the enabling drivers that can affect the test planning and execution process. This includes understanding and clearly communicating the requirements to all participating agencies. Once understood, the test team, through constant communication with the PM, must ensure that the requirements are stable. Failure to keep the baseline requirements in a "box" will predicate changes to the test effort and eventual program development timeline increases. If changes are required, they must be evaluated to determine the impact to the test program. Working within the evolutionary acquisition system, the inclusion of an FOT&E phase to support programmed increments and changes provides the least risk.

The DT community must understand that satisfying the design specifications during DT does not ensure OT success. Understanding the test methodology in the operational evaluation will provide the DT community insight to effectively test a system against specifications as well as operational requirements. Achieving this is possible by following a few guidelines:

- Involve OT early and throughout the process;
- Communicate / understand / confirm system requirements;
- Translate requirements to test events;
- Communicate DT and OT resource requirements early; and
- Generate and evaluate scenarios in DT at the OT level as defined in the OTD Guidebook.

Many of the test deficiencies researched and identified are a direct result of the basic DoD philosophy and test approach. There is a resistance to learn about a system's capability early in its development. The driving factors, aside from budget, are the immaturity of emerging technologies to support the developing product, and the emphasis placed on passing or failing a test rather than learning.

With the new acquisition policies, and emphasis on evolutionary and integrated testing, there is some promise of change. NAVAIR and COMOPTEVFOR are initiating a movement toward integrated testing. While the outcome and benefit of such an initiative will not be known for some time, other programs should begin to evaluate how they are conducting their test programs and compare and contrast it with the integrated process in order to more efficiently make changes if there are positive results from this pilot test program.

THIS PAGE INTENTIONALLY LEFT BLANK

# IV. APPLYING THE PROCESS TO AARGM

## A. INTRODUCTION

Research to this point has offered insight into DoD test philosophy and the differences from the more successful commercial sector philosophy. In addition, lessons learned from past DoD studies and previous development programs were discussed. This information was presented to increase the knowledge base for an individual who may be tasked with developing the test strategy for a major acquisition program. One such program that is in need of developing a robust and effective strategy is the AARGM Program. Within this chapter, there will be a discussion of the weapon system and the test approach that the team is either taking or should consider.

## B. EVOLVING THREAT

Currently in the United States military weapons arsenal, the HARM is the only air-to-ground anti-radiation weapon deployed on tactical aircraft in support of the SEAD mission. Technology and tactics advancements associated with enemy Integrated Air Defense Systems (IADS) have made the SEAD mission of locating, targeting, and engaging these threats increasingly difficult. As demonstrated in Kosovo and Iraq (Desert Storm and Operation Iraqi Freedom), air defense units are becoming more mobile and are effectively employing countermeasures such as Emissions Control (EMCON), blinking, and shutdown to further complicate the SEAD mission. US force structure dynamics demand efficiency in conducting the SEAD mission to reallocate multi-role aircraft to the strike mission. Because of increased political sensitivities to collateral damage and civilian casualties, more effective target location and discrimination are required. HARM guides towards the emitted radiation of enemy radar systems; however, it cannot autonomously yield the target location/discrimination necessary to meet current Positive Combat Identification (PCID) Rules of Engagement (ROE) for current day operations. Additionally, after launch the HARM provides no definitive indication of weapons effectiveness or location of impact to support re-attack decisions or the Battle Damage Assessment (BDA) process. Current HARM employment and effectiveness continue to be limited by simple enemy tactics and the high potential for collateral damage in operations.

## C. MISSION REQUIREMENT

A number of official documents and forums cite the requirements for increased capabilities for reactive or concurrent Joint Suppression of Enemy Air Defense (J-SEAD). A key document, which highlights these requirements, is the Combat Mission Need Statement (MNS) CAF329-92 for Lethal J-SEAD that calls for the reactive destruction of enemy IADS using on-board and off-board sensors. The AGM-88E Advanced Anti-Radiation Guided Missile (AARGM) further addresses key shortfalls based on the results of the J-SEAD Joint Mission Area Analysis and the Joint Requirements Oversight Council's (JROC) approved Theater and Air Missile Defense (TAMD) MNS. Current HARM shortfalls were also discussed and documented at the Strike Weapons Operational Advisory Group (OAG) in 1998 and 2001, and at the Anti-Radiation Missile (ARM) Steering Committee (ASC) in November 1998. Although the above-mentioned initiatives addressed the J-SEAD needs for reducing the timeline for attack on IADS, it did not address key issues such as responsive re-attack, second sensor confirmation in support of ROE, and rapid and reliable weapons impact assessments as part of the BDA process. These three issues were the genesis of the Quick Bolt (QB) Advanced Concept Technology Demonstration (ACTD). These operational issues and requirements are detailed in the United States European Command (USEUCOM) QB Functional Requirements Document (FRD). The AGM-88E AARGM Operational Requirements Document (ORD) for this new SEAD weapon, which takes into consideration all the above requirements for Time Critical Strike (TCS), ROE, BDA, was approved in June 2003.

## D. BASELINE AARGM PROGRAM

The AGM-88E weapons system, Figure 31, currently in SD&D is an upgrade to the current HARM system. The system builds upon the lessons learned from the AARGM ATD and the USEUCOM-sponsored QB ACTD.

The AARGM ATD initiated development of an enhanced seeker for the existing HARM airframe. These enhancements included:

- Weapon Accuracy—Global Positioning System/Inertial Navigation System (GPS/INS) for mid-course guidance during missile flight to target;
- Autonomous Target Location—Improved Anti-Radiation Homing (ARH) seeker field of view, sensitivity, and direction-finding accuracy with autonomous target detection, identification, tracking, and target ranging;
- Improved Lethality—Active Millimeter Wave (MMW) radar, providing terminal homing even in the presence of emitter shutdown; and
- Reduced Collateral Damage/Fratricide—Inclusion of GPS/INS supports the establishment of geographic boundaries. Aircrew can now prevent weapon from impacting within a region, called an Impact Avoidance Zone (IAZ) or exiting a defined boundary, called an Area of Responsibility (AOR).

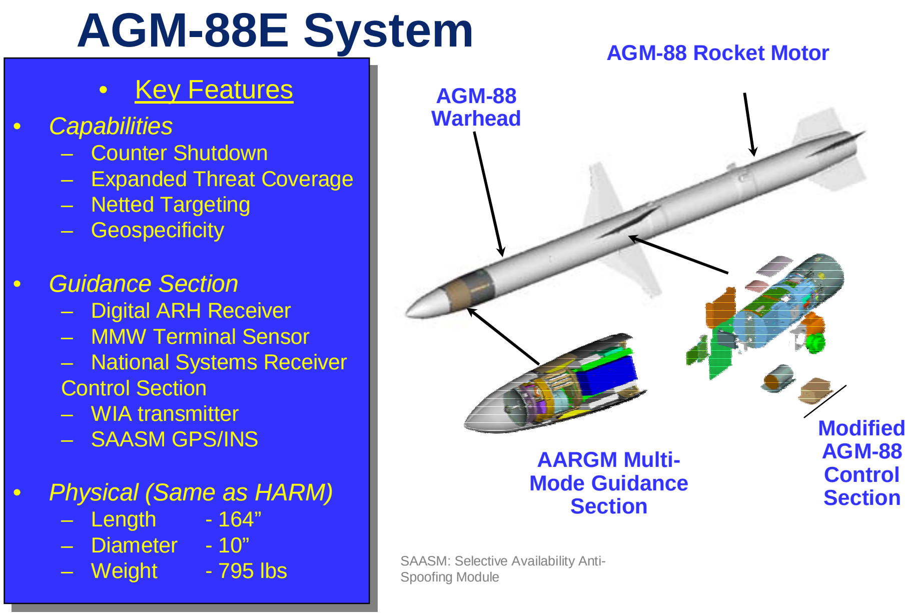


Figure 31. AGM-88E AARGM Missile

Although the AARGM ATD initiative addressed some of the needs identified in the MNS and Fleet forums, it did not address responsive re-attack, second sensor confirmation in support of ROE requirements, and rapid and reliable Weapons Impact Assessment (WIA) indications. The QB ACTD by teaming with the National Reconnaissance Office (NRO) achieved these requirements. Building upon the AARGM ATD, the QB ACTD introduced a major enhancement to the weapon system. This was the inclusion of the national systems architecture. The national systems introduced a net-centric capability to the weapon system and the tactical cockpit. This provided:

- Improved Situational Awareness (SA)—360° reception and display of threat systems provided by the Intelligence Broadcast Service (IBS) and the ARH receiver;

- Improved Targeting—Reception between national systems sensors, ARH receiver, and onboard aircraft sensors enable autonomous multi-source correlation; and

- BDA Support—Missile WIA transmitter injects accurate and timely information in the national architecture supporting BDA for re-attack/combat assessment.

The QB ACTD test program was conducted from November 2002 until September 2003 and included two firings. During the tests, the program successfully demonstrated the capability of transmitting WIA information across national systems to a China Lake strike cell. The information sent by the weapon before impact was received by a ground station and then rebroadcast across the national architecture where it was received by the strike cell. The information was timely, accurate, and supported re-attack decisions or combat assessment.

The second major capability demonstrated was enhanced targeting and increased SA within the tactical cockpit. Achieved by the inclusion of an Embedded National Tactical Receiver (ENTR), the receiver allowed targeting information broadcast through the IBS to enter into the tactical cockpit, thereby providing essential targeting system information. Demonstrated in the last firing scenario, presented in Figure 32, the aircrew, through their cockpit displays, identified, handed-off, and fired upon a correlated target facilitating a successful engagement. This complex scenario included two ambiguous RF targets. The primary target was shutdown while the weapon was in-flight to demonstrate the increased lethality and geographic specificity provided by the weapon system.

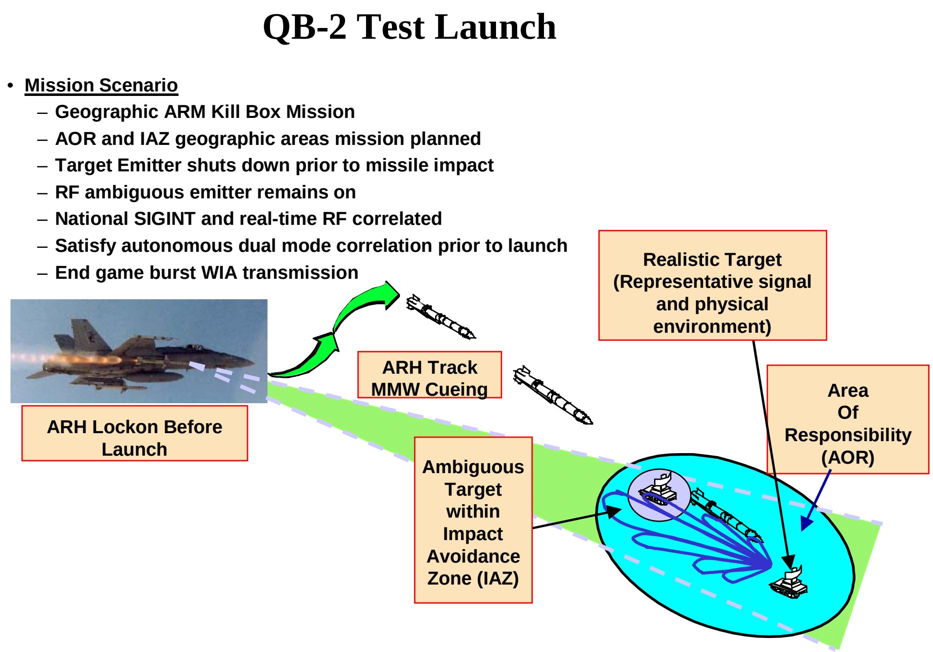


Figure 32. Quick Bolt 2 Firing Scenario

## E. T&E STRATEGY

This advanced and technically challenging development effort is a product of the evolutionary acquisition process. The program entered into the acquisition world as an ATD. The primary focus of this effort was to evaluate the seeker technology and determine if it would be feasible to achieve the desired performance. At the time of inception, the TRL for the enabling technologies each rated a three. (Brady – Deputy AARGM PM, personal communication, July 9, 2004) The key enabling technologies for the program were the advanced ARH receiver and the MMW terminal seeker. The program concluded after five successful firings with an assessment that the technology was feasible. While feasible, there were questions about its producibility and whether the system could support the Sea Power 21 initiative. The program office then entered into an ACTD. During this phase, the maturity of the technology evolved. The contractor also was able to learn about the unique producibility requirements. Because of the combination of the ATD and the ACTD programs, the maturity of the technology at the time of a Milestone B decision improved to a TRL of six.

Currently, according to the deputy program manager, the TRLs for the two technologies are at a seven. (Brady, personal communication, July 9, 2004)

The program's initial test strategy is defined in the TEMP. During the formulation of the TEMP, the test team, composed of members from the program office and the development and operational communities, flowed down the requirements listed in the ORD and translated them into Critical Technical Parameters (CTP) and COIs, as shown in Figure 33. The CTPs are the technical parameters, as defined in the specifications that the DT community will use as their primary metrics. The COIs were carefully developed from the Measures of Effectiveness (MOEs) and Measures of Suitability (MOSs) to represent operational characteristics.

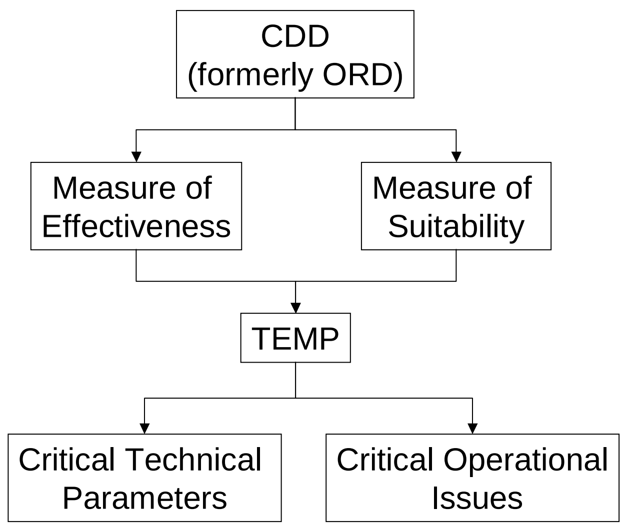


Figure 33. Requirement Flow for TEMP Development

During the flow process, MOE and MOS, defined in the ORD, were evaluated for testability. The test program schedule was developed and recorded in the document. The DT period began in March 2004 and will conclude in March 2008. The DT phases of test are very structured. The first phase is contractor-led and is designated DT-B1. It covers the period from March 2004 until October 2007. During this phase of testing, the primary objective is to develop software algorithms, integrate the hardware specifically located in the guidance section and control section, and perform subsystems testing. The second phase of developmental testing is government-led, covers the period from March 2007 until March 2008, and is designated DT-B2. During DT-B2, the government will test the complete weapon system and the integration between the weapon and the aircraft. The OT community will be given an opportunity to assess the potential suitability and effectiveness of the system during the Operational Assessment (OA), which occurs eight months into the DT-B2 phase. This assessment will be one input into the Low-Rate Initial Production (LRIP) decision. The DT-2B phase will continue throughout the OA. It is the intent of the test strategy to incorporate lessons learned from the OA into the weapon development program. At the completion of DT-B2, the system will enter OPEVAL, which is currently scheduled for June 2008.

With the program more than one year into SD&D, the current policies and plans for the T&E Integrated Product Team (IPT) are beginning to take form. Two areas of focus define the current T&E strategic approach. The first area involves the test planning process, while the second area deals with test execution. The former is critical to the success of the program. To support this effort, the AARGM T&E IPT has created Test Plan Working Groups (TPWG).

> TPWGs facilitate the integration of test requirements and activities through close coordination between the members who represent the material developer, designer, community, logistic community, user, operational tester, and other stakeholders in the system development. The team outlines test needs based on system requirements, directs test design, determines needed analyses for each test, identifies potential users of test results, and provides rapid dissemination of test and evaluation results. (Defense Acquisition University, 2001, p.68)

The AARGM System TPWG includes all stakeholders to the program. They presently meet twice a year to discuss test requirements and the progress made satisfying those requirements. The current membership is follows.

- DOT&E—Provides independent assessments for programs to the Secretary of Defense, USD(AT&L), and Congress.
- N091—Issues policy and procedures for the conduct of Navy T&E.
- COMOPTEVFOR—Navy's sole independent agency for operational test and evaluation.
- VX-9—Operational Test Squadron.
- VX-31—Developmental Test Squadron.
- ARM Weapons Office—Developmental Engineering Group.
- PMA-242—Program Office.
- ATK Missile Systems Corporation—Primary Contractor.
- Joint Interoperability Test Command (JITC)—Command ensures Interoperability KPPs are satisfied.
- National Reconnaissance Office (NRO)—Provides the technology to support National Targeting.
- National Security Agency (NSA)—Provides the necessary protocol to use the national systems.
- F/A-18 Advanced Weapons Lab (AWL)—Develops the aircraft interface software.
- Range Support—Responsible for targets, data collection, instrumentation, and range airspace.

There are also lower echelon TPWGs that concentrate on specific areas to support test. These TPWGs have a reduced membership and focus on developing strategies to overcome specific risks such as target and range limitations or asset utilization. The AARGM TPWGs are currently addressing the following areas in an attempt to establish a plan for a successful T&E program.

### 1. Targets

The terminal seeker for the AARGM system presents some unique challenges for target requirements. The upgraded anti-radiation missile requires two primary features from the targets. The first is a valid RF signal for the ARH receiver and the other is operationally representative Radar Cross Sections (RCS). While the former is a legacy HARM requirement and is typically producible at current test ranges, the latter requirement creates some difficulty. DOT&E sited in their FY03 report on AARGM that there are not enough procured targets within the range infrastructure to support the needs of the AARGM program. (DOT&E, 2004, p. 123) In an effort to address this challenge, the T&E IPT established a Targets Working Group (TWG) dedicated to target support. Their primary responsibility is to evaluate the current asset availability within the existing range infrastructure and develop a strategy to expand it in order to support the AARGM test effort. In the short existence of the working group, they have created a matrix of all available threat systems within the US range infrastructure. This list, while not complete, identifies the location of the system, type (real or simulated), and operational status. Additionally, in an effort to meet some target requirements defined in the ORD/TEMP, as well as those requested by the OT community, they have let small contracts with research universities and NAVAIR range departments, specifically at China Lake, to begin the process of repairing and in some cases developing the threat systems. They have also evaluated the use of overseas ranges. Engaging the operational community early, to solidify their target needs and requirements, has allowed the TWG and the T&E IPT to identify commonality between the contractor, DT, and OT communities' needs. This will allow the test team to more efficiently use the limited target resources. Learning from the ATACMS test effort, the TWG will be identifying the requirements to verify, validate, accredit, and instrument select targets. This constant dialogue between the contractor, DT, and OT communities will help prevent late test target requirements from delaying test execution. While it is necessary to accredit the OT targets, the current plan is to accredit all test targets, thereby increasing the possibility for the OT community to leverage some DT testing with their own.

### 2. Range

Testing against a variety of background scenarios will be required to demonstrate the increased system lethality. This is difficult since most ranges that contain simulated or realistic Surface-to-Air Missile (SAM) systems are located in a desert environment. While searching for acceptable targets, the T&E IPT has been concurrently evaluating possible locations to create operationally representative environments in unrepresentative test areas. Current policy, based on safety concerns, necessitates firing this weapon within a very limited number of test ranges. These ranges are China Lake, Utah Test Range, and the two sea ranges on the east and west coast. A fifth range, located in Roosevelt Road, Puerto Rico, closed recently due to civilian encroachment. As a result of the available ranges, the background environments are limited to desert and sea. As stated in the ORD and TEMP, the weapon system will need to be tested and evaluated against other operationally representative environments to verify system performance. The Range TPWG is currently working with the Targets TPWG to identify possible alternatives to the current challenge. Similar to the targets challenge, consideration is being given to use allied range resources. Other proposals are to augment the target environment at the desert ranges to reflect the other background environments. Each consideration brings challenges. Regardless of the approach, these two TPWGs will need to continue to actively involve the OT community as well as DOT&E to ensure that the background scenarios will be operationally acceptable.

### 3. Personnel

DSB sited in their 1999 study that rotating personnel within the Test and Evaluation organization is a contributing factor to DoD's poor T&E performance. Their recommendation,

> Establish a stable team made up of users, developers, testers and appropriate contractors called a Combined Acquisition Force (CAF) to streamline the acquisition process for ACAT I programs. The CAF should be formed once a need is identified and remain in place throughout the acquisition process. (DSB, 1999, p.4)

Although the AARGM program is an ACAT 1C program, it does not have a CAF. The program does contain some very experienced testers, specifically within the government. They are also well within the retirement age. The program is currently financially limited and does not have the ability to bring in young government test engineers to mentor. This has created a program concern that the loss of key test personnel before the completion of the program will adversely affect the test effort. Recognizing this concern, the T&E IPT focuses on documenting all decisions, processes, and results in an official configuration managed process. Recording the who, what, when, where, why, and how of a decision or event, and correctly archiving it for others to view will help minimize the disruption that is inevitable as the personnel within the T&E team change. Through this effort, mitigating the possibility of not addressing historical deficiencies, similar to the SLAM-ER program, is possible.

#### 4. Operational Involvement

Building upon the recommendations from past studies and lessons from other programs, the AARGM T&E IPT has actively pursued the early involvement of the operational community. This process began during the ATD and ACTD programs. The current operational organizations that have been directly involved in the test planning process since the first TPWG, which began the development process for the TEMP, were DOT&E, N091, VX-9 and COMOPTEVFOR. VX-9 is responsible for executing the operational test, while COMOPTEVFOR acts as the policy manager and ensures the necessary planning and documentation are in place. Current successes for the AARGM T&E effort as a result of working with the operational community early:

- Redefining the KPPs/MOEs/MOSs to ensure testability; earlier definitions as written in the ORD were either ambiguous or not testable;
- Resolution on the number of live firings and the number of live warhead shots;
- Establishment of a dedicated operational assessment during the developmental test period;
- Establishment of three DT assist test phases;
- Establishment of the weapon instrumentation requirements during OT; and
- Understanding of the operational test and financial resource requirements during the OA and the OPEVAL.

Open and continuous communication ensured the establishment of positive relationships and the understanding of various test requirements levied by the operational community. Although successful to date, there are other test issues requiring definition and direction. The involvement of the entire OT community is essential to ensure effective testing.

One such issue is that firing scenarios must be generated for both DT and OT. DOT&E had stated the number of test firings available is insufficient to support the test effort. (DOT&E, 2004, p.123) To overcome this risk, the test team must clearly define the requirements that are to be tested during the firings. To do so they must use dendritics.

> Dendritics is a tool to develop and see relationships. The process of creating the dendritics facilitates the identification of critical issues, Measures of Effectiveness, Measures of Performance, and data requirements. The data requirements then facilitate developing the test plan for a system. By identifying the data requirements necessary to answer the questions posed in the dendritics, testers can formulate tests to capture the necessary data. (Helm, 2002, p.8)

Early involvement by the operational community affords the DT community an opportunity to reflect on environmental considerations during test events.

> When programs do poorly in operational tests, frequently it is because they permit themselves to encounter for the first time some operational environment or requirement that they have never tried before, or have tried before in developmental testing, but only unsuccessfully. This can include environments like rain, dirt, dust, or wind; or it can be countermeasures, realistic threats, or realistic operational environments. For example, the Army's SADARM (Sense and Destroy Armament/Armor) program was doing fine in developmental tests in the clean desert at Yuma Proving Ground, but when they got into the operational test with interesting terrain, trees, and realistic countermeasures, they didn't do so well. (Coyle, 2000, p.3)

The early involvement of other agencies and commands affords the opportunity to address unique planning requirements. The net-centric enabling technologies require the involvement of the NRO, NSA, and JITC. These agencies have requirements not typically considered in basic weapons programs:

- National scheduling
- National targeting information
- Data requirements

Another issue is the data flow chain. Establishing agreement on the data sharing throughout the test process is essential. Leveraging the IT&E concept will afford increased opportunities to share data between the DT and OT communities.

Delivery times for production-representative missiles are another issue. In addition to the delivery times, the DT community is currently working with DOT&E to clearly establish the definition of production-representative systems. During the early TEMP development efforts, there was a discrepancy between the two agencies' interpretations. By engaging the DOT&E early, there has been time to develop a strategy. At the time of this research, the strategy proposed has not been officially accepted by DOT&E.

Another concern is the agreement on the TM section requirements for the OT community. Currently there are two versions for the weapon's telemetry sections. One provides a higher fidelity of data but incorporates a non-production representative filter in the weapon. The other provides less information but maintains the integrity of the production-representative configuration. The DT community is currently working with the OT community as well as DOT&E to maximize the use of the higher data rate TM sections during OT. This will offer increased system performance knowledge throughout the test effort.

Establishing an integrated T&E effort that will more efficiently use the limited range and financial resources is also desirable. One proposal recommends removing one firing from the OA. The reason is that, based on the current contractor delivered software schedule, the release of full functionality software occurs just before the start of the OA. This will not afford the DT community time to conduct the necessary preliminary tests thereby increasing risk for a successful outcome from the OA. With the immature software, there would be limited knowledge gained from a second firing. Allocating it into a later integrated test firing with more mature hardware and software offers an increased opportunity to learn more about the system performance.

The creation of the developmental test scenarios is benefiting from the early involvement of the operational community and its inputs. At a recent operator's (user) meeting, the AARGM IPT lead requested that the Fleet subject matter experts send training scenarios that include the use of ARM weapons. The intent is to use these Fleet training scenarios as a foundation for the DT firing scenarios. Scripting tests the way the user will fight with the article offers a plethora of potential knowledge about the system's maturity. Where an operational scenario cannot support a live-fire event, captive testing will be performed.

### 5. Managing Requirements

Requirement stability is an important aspect for the establishment of a secure test strategy. "Change in requirements was identified as a major problem for T&E…difficulties in defining test requirements made test planning and the conduct of tests more difficult and expensive than originally estimated." (Hoivik, 2000, p. 36) The AARGM T&E IPT continues to face emerging requirements to demonstrate increased capability. The program, originally divided into three evolutionary phases, was combined into one phase. In the original acquisition strategy, the WIA capability and the national targeting capability were product improvement initiatives. As program pressure to provide increased capability grew during the Milestone B decision, the program was re-scoped and the phases combined as a baseline capability. This decision increased the focus of test and evaluation without the benefit of time or funding. Complicating the situation, the re-scope decision was made without the involvement of the test team.

Attempting to minimize requirements creep, the T&E IPT is increasing their dialogue with the PM. This has offered opportunities to express concerns about funding and schedule, when additional system capability is being considered. The team is also working with the contractor's systems engineering team. The systems engineering team is using the DOORS® engineering tool to flow and track operational and technical requirements. DOORS® is a requirements management tool designed to capture, link, trace, analyze, and manage a wide range of information to ensure a project's compliance to specified requirements and standards. (Telelogic home page, retrieved August 29, 2004) A by-product of this tool is a matrix that can be used by the T&E IPT to develop a test point matrix. The value that this provides is a clear relationship path to a test event and the requirement. If requirements are added without the T&E IPT's knowledge, it will be reflected in the computer generated matrix. This tool also will allow the T&E IPT to clearly define when a specification/requirement is being tested and by what agency (i.e., contractor or government DT). The AARGM test team is adopting the lessons learned from the F/A-18E/F test program. This program was successful because of their strict adherence to the baseline requirements during the initial development and test effort.

### 6. PM and Tester Relationship

The distrust between the PM and the tester that has been identified in a variety of references and sited in this research will consciously be avoided. While different personalities will migrate into the PM and testing community throughout the effort, open and honest communication has proven and will continue to prove effective. Additionally, averting or minimizing conflict is possible if the T&E IPT follows adherence to recording decisions and accurately tracking actions between the two groups. The T&E IPT can further minimize conflict by identifying very early the desired test schedule and objectives for each of the test events. Establishing the entry and exit criteria for test events early in the planning effort, and getting PM approval, will provide the T&E IPT a solid foundation to work from throughout the test phase especially during time sensitive test periods. This process is currently beginning within the T&E IPT, as they begin to define the objectives for the DT firings and the scope of the OA.

### 7. Suitability

The AARGM T&E IPT, through the Integrated Logistics Support (ILS) IPT is actively pursuing any areas that could present difficulty during the T&E phase. Historically operational suitability has proven to be a source of program failures during the OPEVAL phase of test. "The Army has seen that 80% of their systems have not met 50% of their reliability requirements in operational test." (Umansky, 2001, ¶9) Suitability encompasses a variety of areas, which are evaluated by the operational community. With the requirements clearly defined within the OT test guide and the TEMP, the DT community has actively pursued a roadmap to ensure compliance. Key areas of interest include aircrew and maintenance training and support, reliability, and maintainability. These have been areas of weakness for previous HARM development efforts. As a result, they are receiving increased attention early in the development phase.

Currently the ILS team has conducted meetings with operational maintainers and users to identify concerns. Application of lessons learned from the AIM-9X program will prove beneficial. As stated earlier in the research, the AIM-9X program identified major shortfalls early in the development phase by incorporating the OT community. Active utilization of ILS modeling tools, such as the NAVAIR-developed Audit Trail, has recently identified a requirement discrepancy between the system specifications and the ORD. (Chapman, 2004) As a result, a recommendation to modify the system specification has been requested to the PM.

## F. SUMMARY

The AARGM TEMP, officially signed by DOT&E on August 12, 2004, clearly states the challenges faced by the T&E IPT.

> We reviewed and subsequently approve the attached AARGM TEMP No. 1651, dated June 10, 2004. This is a success oriented test program; however, performance shortfalls may require additional test assets to ensure an adequate test and the successful execution of operational mission scenarios. (OSD Memorandum, 2004)

Added to those comments are the program risks previously recognized by DOT&E.

- Test range infrastructure does not exist to adequately assess the full capabilities of the design with regard to target discrimination.
- Limited number of missiles available during testing.

(DOT&E, 2004, p.125)

These are realistic challenges, sited by an organization that has a holistic view of all of DoD test programs. While challenging, they are surmountable. The AARGM test strategy is based on a careful assessment of:

- Current AARGM program requirements;

- Involvement of all agencies related to weapon development, test, and use;
- Understanding of available range resources; and
- Understanding of DoD program lessons learned.

These assessments are based on the research that was used to support this thesis. Table 1 presents the challenges identified within this research, and the current mitigation strategy adopted by the T&E IPT to address those challenges.

Table 1. Strategy to Overcome AARGM Challenges

| AARGM T&E Element | Mitigation Strategy | Some Research Support Areas | Section of Thesis |
|---|---|---|---|
| Targets | -Current Resource Availability (US and Foreign)<br>-Integrated Test Team<br>-Targets Working Group | -DOT&E Study, AIM-9X Program | II |
| Range | -Evaluation of Range Complexes (US and Foreign) | -BRAC, Range Encroachment, DOT&E Study | II |
| Personnel | -Documentation and configuration control of Decisions, Processes and Results | -Commercial Philosophy<br>-DOT&E Report, DSB Study | II |
| Operational Involvement | -Early Involvement at ATD / ACTD Level<br>-Inclusion of OT and DOT&E in original test planning process<br>-Test Plan Working Group<br>-Integrated Test Team | -Several Past Studies and Findings<br>-DSB & SAIC Studies, NAVAIR Study<br>-SLAM-ER, Army Cargo Trailer | I, II, III |
| Managing Requirements - System and Test | -Integrated Systems Engineering Team<br>-Integrated Test Team (OT/DT/DOT&E)<br>-Communication<br>-Use of commercial systems engineering tool | -Several Past Studies and Findings<br>-F/A-18E/F, F/A-22, ATACMS Program<br>-Boeing Lesson Learned | II, III |
| PM and Tester Relationship | -Communication<br>-Establishing Exit and Entry Criteria<br>-Documentation of Decisions | -Commercial Test Philosophy<br>-DSB Study, SLAM-ER Program | II, III |
| Suitability | -ILS Modeling Tools / Audit Trail<br>-Early OT Involvement | -AIM-9X Program<br>-NAVAIR Study<br>-Tactical Tomahawk | III |

Awareness of the challenges will not ensure success, but it will afford the test team an opportunity to reduce risk to an acceptable level.

# V. CONCLUSION

## A. RECOMMENDATIONS

DoD is developing CONOPS for the future battlefield that demands the procurement of technologically advanced and highly integrated systems to support the Warfighter. These systems, in development or in a conceptual phase, will require a product development approach that ensures they end up in operational use, on schedule, within cost and meeting all performance objectives. This approach is the evolutionary acquisition approach. “Evolutionary acquisition is an approach that delivers capability in increments, recognizing, up front, the need for future capability improvements. The objective is to balance needs and available capability with resources, and to put capability into the hands of the user quickly.” (DoDI 5000.2, 2003, p.3) With this approach, T&E faces new challenges in their mission. It:

- Requires more flexible test planning to deal with undefined thresholds;
- May require more testing to insure no adverse effects on earlier capabilities;
- Complicates logistical support and evaluation of suitability;
- Requires constant coordination between user, developer, and testers; and
- Overall cost of test may go up.

(Lockhart, 2002)

Because of these new challenges, PMs must embrace the important role that T&E plays within their program. Recognizing that T&E is more than a single phase on the development schedule will enhance the product development process.

> Testing is an essential component of the systems engineering processes. Too often, DoD has viewed testing as a disconnected single event or milestone through which systems must pass. Testing should be a process that begins on day one and continues throughout the design life of any system. This is especially true when one considers the new evolutionary acquisition model. This model embraces the concept of spiral development and encourages rapid technology insertion. In this model, testing is critical to producing and improving overall systems by integrating knowledge about the impact of each technology insertion into the development cycle. (Sega, 2003, p. 7)

Testers further should recognize that they now play a larger role in the process and work to establish the necessary processes to effectively contribute to the weapon system development. There are five principles that the tester and the PM should accept as they embark on this teaming venture. They should:

- Develop meaningful and applicable test objectives, and adhere to them in an orderly, repeatable, and disciplined manner;
- Use the closed loop systems engineering approach, from concept, to component, to subassembly, to subsystem, to system, to whole system test;
- Test as early as possible and as often as affordable to find and correct problems before they become too costly;
- Involve the user, developmental tester, and operational tester in the initial formation of the systems engineering council to develop test objectives to ensure continuous and timely information exchange of objectives and test results; and
- Take the time to ensure all parties (developer, contractor, and government operational testers) thoroughly understand the system mission requirements and agree on how the system will be tested, scored and evaluated.

(Bodmer, 2003, p.68)

Recognizing the role of T&E and acceptance and adherence to principles designed to improve it will mitigate the current trend seen in each Services poor operational test results. Other areas that the DT&E community can improve upon to strengthen weapon system development include:

- Range Infrastructure Capabilities – A Warfighter does not enter a battle without clearly understanding the battlefield. To do so would lead to defeat. A tester must understand the range resource environment, which includes ranges, capabilities, and personnel, and effectively use what is available and quickly highlight the limitations. Understanding the limitations early will afford time to develop alternative methods of test;
- T&E Training – Providing the necessary tools for a successful T&E effort starts with training the workforce. Tester and PMs must be adequately trained in the field of T&E to understand the challenges that they will face. This training should also afford them the knowledge of past program efforts; and
- Integrated T&E Efforts – Stovepipe approaches to T&E do not foster a successful program. The DT community must actively pursue the involvement of the OT community early in the planning of a test program. Understanding the testing needs of the OT community and validating decisions made by the PM or the DT community will reduce expending limited resources. In addition, it will identify any conflicts early and afford time for resolution without affecting the program schedule. DoD must globally recognize and fully support the integration of test. With limited resources, contractor, DT, and OT test phases should leverage from one another to reduce repetition.

No contractor involvement in the operational test phase will hinder acquisition streamlining, because the recovery period after the test will be made longer. The contractor will have to wait until the end of the test before any fixes can be applied and tested. This will

> make the total test time longer and more expensive. The total acquisition period will also be longer, again raising total program cost." (Stoddart, 2001, p.5)

The commercial industry has learned a great deal about effectively testing a system. Folding their experiences into the DoD test process will enhance government test efforts. These include:

- Knowledge based test approach – The concept of testing a developing system to a high level of fidelity early will offer keen insight into the maturity of the system. The DT community must resist the urge to delay complex testing until later in a program's product development schedule for fear of failure. While the commercial sector has supported this approach because of lessons learned, DoD has not. With the complexity of systems increasing, this concept will become the distinction between successful and unsuccessful programs;
- PM, tester, and contractor relationship – The test team must promote effective communication among the various organizations involved in the test process. Program foundations built upon positive communication will reduce the negative relationship between the tester and the PM. This approach will promote aggressively handling problems earlier in the test cycle; and
- Lessons-learned forum – DoD does not offer a means to easily learn lessons from other program efforts. While some information is available, it requires a dedicated effort, like thesis research, to gather the data. DoD needs to consider establishing an improved forum to distribute T&E lessons learned.

While the above recommendations are global, they also can afford the AGM-88E program guidance to effectively test the system during the development phase. Practices that have already been established within this program include the early involvement of the operational test community, positive communication with the PM and contractor team, and an identification of the range resource limitations. While these efforts have ensured early resolution or a path ahead for currently identified T&E concerns, other practices require adoption:

- Define the integrated testing methodology and scope;
- Establish Operational Test requirements for targets and instrumentation;
- PM acceptance to accelerate the complexity of scenario-based testing;
- Resistance to accept any new system requirements during the test phase; and
- Establish an internal T&E training program to include T&E lessons learned from other programs.

The top four recommendations, when complete, require recording in the TEMP. This will ensure that all involved in the program's development effort understand the current test strategy.

## B. PROPOSED FURTHER STUDY

This research was originally designed to discuss the AGM-88E T&E effort and identify ways to ensure success during OT. During the research phase, it became apparent that the scope was going to expand in order to understand the current DoD situation with testing. It also became apparent that the entire problem could not be fully evaluated. As a result, there is a variety of follow-on research possibilities:

- Evaluate the impact that effective training can have on the T&E community. With the continuously changing acquisition environment, it is imperative that the workforce understands the documentation and practices that support the T&E effort as well as the lessons from past programs;

- Analyze the Integrated T&E Process and identify how this approach to testing will influence the acquisition process. COMOPTEVFOR and VX-9 are the driving forces behind the integrated test approach within naval aviation acquisition. As this approach is new, T&E documentation does not reflect the process. The Navy is going to implement this concept into a pilot program associated with a software development effort for the F/A-18. An analysis of the program's performance and the key enabling concepts for the IT&E effort would prove beneficial to future programs; and
- Evaluate the practices and processes used by the AARGM T&E IPT. Faced with many challenges, explore the direction the program has ventured and evaluate if the accepted practices have resulted in a positive T&E program.

## C. CONCLUDING COMMENTS

Testing a developing system in DoD can be a challenging and rewarding experience. The key to success is to understand the system, the operational environment, and the lessons learned from others who have come before. Failing to recognize the importance of the latter will lead to repeating similar mistakes resulting in inefficiency and possible program cancellation. The DT community plays a tremendous role in the success of the program. They must fully understand the program and its requirements. Their test planning strategies will be the basis of evaluating the product before going into operational test. If they should fail to effectively identify performance or suitability issues, the chances for success decline. The challenges facing the AARGM T&E IPT team are tremendous. The AGM-88E weapon system is a vast improvement over the current SEAD system. With this improvement, there are increased T&E demands. There are many challenges, and the T&E team must clearly identify and communicate them to the PM. In addition, they must continue to refine their test strategy to ensure executing in the most effective and efficient manner. The

Warfighter is expecting to have the capability in 2009, and it is the responsibility of the AARGM Team to deliver the product on time and with all performance objectives achieved.

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF REFERENCES


AIR 5.1E Acquisition T&E Department Brief, Operational Test Success: Lessons Learned Assessment, March 11, 2004

AIR 1.6 Brief to AIR 1.0, OT Success: Lessons Learned Assessment, January 5, 2000

Air Test and Evaluation Squadron Three One (VX-31) Third Quarter FY04 Assessment, July 2004

Alford, Lionel LtCol (USAF). "How Compensation in Test and Evaluation Affects Aircraft Acquisition," *Defense Acquisition Review Journal* (January – April 2004)

Battershall, Lee. "Technology Approach: DoD Versus Boeing (A Comparative Study," *Acquisition Review Quarterly* (Summer 1995)

Bennitt, Brent VADM (ret). "Building Test and Evaluation Transformation on a Solid Foundation," *ITEA Journal* (December 2003 – January 2004)

Bodmer, Gerald. *Test and Evaluation in the United States Navy, and How It Must Evolve to Support Future Systems Acquisition*, Master's Thesis, Monterey, CA. The Naval Postgraduate School, September 2003

Burris, Stephen CAPT (USN). Former Commanding Officer VX-31, interview, June 2, 2004

Chapman, Thomas. "Overview of Reliability, Availability, BIT Diagnostics, T&E Requirements on AARGM," White Paper, August 2004

Christie, Thomas. "Test and Evaluation in the New World," presented to National Defense Industrial Association, March 2002

Christie, Thomas. "Test and Evaluation in the New World of 2004," presented to National Defense Industrial Association, March 2004

Converse, Mark CAPT (USN). PMA-242, interview, June 7, 2004

Coyle, Philip III. Evolutionary Acquisition: Seven Ways to Know if You are Placing Your Program at Unnecessary Risk. *Program Manager*, (November/December 2000)

Cotterman, Howard, Kevin Forsberg and Hal Mooz. *Visualizing Project Management: A Model for Business and Technical Success*. New York: John Wiley & Sons, Inc, 2000

COMOPTEVFORINST 3960.1, Operational Test Guide, January 2004

Daly, Stephen, Larry Miller, Dr. Ernest Seglie, and George Wauer. "The Test and Evaluation Master Plan: An Operational View." *ITEA Journal of Test and Evaluation* 24/3 (September/October 2003)

Defense Acquisition University. Systems Engineering Fundamentals, January 2001

Defense Science Board Task Force on Test and Evaluation, September 1999

Dillard, John T. "IOT&E from the PM's Perspective," Army Tactical Missile System After Action Report, June 27, 1990

Dillard, John T. "Preparation for Operational Testing," draft white paper, August 2003

DoDD 3200.11, Major Range and Test Facility Base, December 1, 2003

DoDD 5000.2, The Defense Acquisition System, May 2003

DoDI 5000.2, Operation of the Defense Acquisition System, May 2003

DOT&E FY01 Annual Report, February 22, 2002

DOT&E FY02 Annual Report, March 5, 2003

DOT&E FY03 Annual Report, January 20, 2004

Director, Operational Test & Evaluation, "Test and Evaluation's Role in Experimentation," accessed August 9, 2004 from http://www.dote.osd.mil/presentations/paxspeech.html

Dulin, Patrick. "Issues and Challenges Facing the T&E community with Rapid Weapon System Technology Changes," *Program Manager* (November – December 2001)

F-22 Raptor article accessed June 5, 2004 at http://www.globalsecurity.org/military/library/budget/fy2001/dot-e/airforce/01f22.htm

Federation of American Scientists. Tri-Service Standoff Attack Missile article accessed June 5, 2004 at http://www.fas.org/man/dod-101/sys/smart/tssam.htm

GAO Report, *A More Constructive Test Approach is Key to Better Weapon System Outcomes*, GAO/NSIAD-00-199, July 2000

GAO Report, *Actions Being Taken to Address Testing Recommendations, but Updated Assessment Needed*, GAO/NSIAD-04-254, February 2004

GAO Report, *Army Purchased Truck Trailers that Cannot Be Used as Planned*, GAO/NSIAD-00-15, October 1999

GAO Report, *Assessments of Major Weapons Programs*, GAO/NSIAD-03-476, May 2003

GAO Report, *Better Management of Technology Development Can Improve Weapon System Outcomes*, GAO/NSIAD-99-162, July 1999

GAO Report, *Better Matching of Needs and Resources will Lead to Better Weapon System Outcomes*, GAO-01-288, March 2001

GAO Report, *DoD Training Can Do More to Help Weapon System Programs Implement Best Practices*, GAO/NSIAD-99-206, August 1999

GAO Report, *Improved Program Outcomes Are Possible*, GAO/T-NSIAD, March 1998

GAO Report, *Readiness of the Marine Corps' V-22 Aircraft for Full-Rate Production*, GAO-01-369R, February 2001

GAO Report, *Status of the F/A-22 and Joint Strike Fighter Programs*, GAO-04-597T, March 25, 2004

Gehrig, John, Gary Holloway, and George Schroeter. “Reflections on Test and Evaluation: T&E Infrastructure, Reengineering Army T&E, and Building a Viable Test Range Complex – T&E – DoD Test and Evaluation,” *Program Manager* (July-August 2002)

Helm, Eric. *Operational Scenarios for the Initial Operational Test and Evaluation of the RAH-66 Helicopter*, Master’s Thesis, Monterey, CA. The Naval Postgraduate School, June 2002.

Hoivik, Thomas. OS-4601 Lesson Lecture, Naval Postgraduate School, May 1997

Hoivik, Thomas. “Recurring Lessons in Weapons T&E Programs,” *Army AT&L* (September – October 2000)

Institute for Defense Analysis. “New Initiatives for Operational Test and Evaluation,” *IDA Research Summaries* 6/2 (Fall 1999)

Johnson, Collie. “NDIA National Summit on Acquisition, Research, Test and Evaluation,” *Program Manager* (July – August 2001)

Levin, Bob. “Model for Knowledge-Based Acquisition Process,” presented at the International Test & Evaluation Summit, February 25, 2003

Lockhart, Richard. “Integrating Test and Training,” presented to SETE 2002 Conference, October 2002

Office of the Secretary of Defense. “Approval of the Advanced Anti-Radiation Guided Missile Test and Evaluation Plan No. 1651,” Memorandum for ASN RD&A, August 12, 2004

Owen, Walter, DPA, MN4602 Lesson Lecture, Naval Postgraduate School, February 2004

Pike, John. A-12 Avenger II article accessed June 18, 2004 at http://www.globalsecurity.org/military/systems/aircraft/a-12.htm

Rogers, Edward Ph.D. and Col. Robert Birmingham, USA (Ret), "A Ten-Year Review of the Vision for Transforming The Defense Acquisition System," *Defense Acquisition Review Journal* (January – April 2004)

Science Applications International Corporation, *A Study of Commercial Industry Best Practices in Test & Evaluation Which are Potentially Applicable to DoD Developmental Test and Evaluation*, July 2002

Sega, Ronald Dr. "Research, Development, Test and Evaluation," *ITEA Journal* (June – July 2003)

Stoddart, John Col USA (ret). "Contractor and Operational Testing: Some Involvement is Legal and Beneficial," *Program Manager* (March – April 2001)

Telelogic web page, DOORS® definition, accessed August 29, 2004 at http://www.telelogic.com/products/doorsers/doors/index.cfm

Umansky, Eric. Studs and Duds. *Washington Monthly* (December 2001), accessed December 13, 2001 at http://www.washingtonmonthly.com/features/2001/0112.umansky.html

Ward, Dan Capt. "Modern Acquisition Myths: One Size Does Not Fit All," *Program Manager* (March – April 2003)

Wascavage, Joe. "Naval Air Systems Command and Test and Evaluation," OTD Brief, March 2004

Web article, GAO Report Large Cost Overruns and Lack of Testing and Accountability for Missile Defense Programs, accessed July 8, 2004 at http://reed.senate.gov/press108th/Defense/rel-gao-missile-defense-04-23-04.htm, July 8, 2004

Wynne, Michael W. Testimony before the Senate Armed Services Committee, May 21, 2002, retrieved from http://www.norad.mil/index.cfm?fuseaction=home.news_rel_03_31_04, June 10, 2004

THIS PAGE INTENTIONALLY LEFT BLANK

# APPENDIX

| Technology readiness level | Description |
|---|---|
| 1. Basic principles observed and reported. | Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a technology's basic properties. |
| 2. Technology concept and/or application formulated. | Invention begins. Once basic principles are observed, practical applications can be invented. The application is speculative and there is no proof or detailed analysis to support the assumption. Examples are still limited to paper studies. |
| 3. Analytical and experimental critical function and/or characteristic proof of concept. | Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative. |
| 4. Component and/or breadboard validation in laboratory environment. | Basic technological components are integrated to establish that the pieces will work together. This is relatively "low fidelity" compared to the eventual system. Examples include integration of "ad hoc" hardware in a laboratory. |
| 5. Component and/or breadboard validation in relevant environment. | Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so that the technology can be tested in a simulated environment. Examples include "high fidelity" laboratory integration of components. |
| 6. System/subsystem model or prototype demonstration in a relevant environment. | Representative model or prototype system, which is well beyond the breadboard tested for TRL 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high fidelity laboratory environment or in simulated operational environment. |
| 7. System prototype demonstration in an operational environment. | Prototype near or at planned operational system. Represents a major step up from TRL 6, requiring the demonstration of an actual system prototype in an operational environment, such as in an aircraft, vehicle or space. Examples include testing the prototype in a test bed aircraft. |
| 8. Actual system completed and "flight qualified" through test and demonstration. | Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications. |
| 9. Actual system "flight proven" through successful mission operations. | Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. In almost all cases, this is the end of the last "bug fixing" aspects of true system development. Examples include using the system under operational mission conditions. |

THIS PAGE INTENTIONALLY LEFT BLANK

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California